

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 27

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5943 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) IMMEDIATE SOURCE:

(B) CLONE: Human Fibrocytes A-alpha chain

(ix) FEATURE:

- (A) NAME/KEY: CD6
- (B) LOCATION: join3 L84, 1154..1279, 1739..1922, 3055..3200, 3786..5210)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTCTAGGAAC CAGCCCCACC CTAGAAAAAG ATG TTT TCC ATG AAG ATC GTC TGC	54
	Met Phe Ser Met Arg Ile Val Cys
	1 5
CTA GTT CTA AGT GTG GTG GGC ACA GCA TGG GTATGGCCCT TTTCATTTTT	104
Leu Val Leu Ser Val Val Gly Thr Ala Trp	
	10 15
TCCTCTGCT TTCTCTCTGG TGTTTATTCC ACAAAGAGCC TGAAGGTCAO AGTCTACCTO	164
CTCTATGTCC TGACACACTC TTAGCTTTAT GACCCCAAGC CTGGAAGGAA ATTTCCTGGG	224

-continued

TOGGCTTOAC	ACCTCAAGAA	TACAGGOTAA	TATGACACCA	AGAGGAGAT	CTTAGATGAA	284
TOGAGGTOTA	CAACTACAAO	GGAAACTTTA	GCATCTOTCA	TTCAGTCTTA	CCACATTTTO	344
TTTTGTITTO	TTTTAAAAAO	GGCAAGAAAT	ATTTOCCATC	CTTGTACCTA	TAAAGCCCTO	404
GTGCATTATA	ATGCTAGTTA	ATGGAATAAA	ACATTTTATO	GTAGAGATTTO	TTTTCTTTAG	464
TTATTAAATTT	CTTGTACTTT	GTCCATAATA	AGCAGAACTT	TTAGTGTTAG	TACAGTTTTO	524
CTGAAAGGTT	ATTGTTGTTT	TTGTCAAGAC	AGAGAGAAAA	GCAACGGAAT	TATCTTTTGA	584
AATATCTTTO	CAGTATCAGA	AGAGATTAGT	TAOTAAAGCA	ATACGCTTTT	CCGAGATAAT	644
GGTATTCTTT	TAAATTATGA	ATCCATCTCT	AAAAGTTTACA	TAGAAACTTO	AAAGGAGAGG	704
GAACATTCAO	TAAAGATAAT	CTAGGTTTTT	CTACTGAAAC	AGCAATTACA	GGAGAGAGAG	764
CTCTACAGTA	GTTTTCAACT	TCTGTCTGOC	AGTCATTAGT	AAAAATGAAA	AGGTAAAAAT	824
TAACTGATTT	TATAGATTCA	AATAATTTTT	CTTTTAGGAT	GGATTCTTTA	AAACTCTCTA	884
TATTTATCAA	ATGCTTATTT	AAOTGTACCA	CACAGTTTAA	AAATTTGTAC	ACCTTGTCTC	944
CTTAAATTTCT	CATAACAACCT	CCATAAAATO	GGTCTTAAGA	TTTCCATTTT	AAAGTAAAGAA	1004
ACCTGAAAGCT	TGCCGAAAGCC	CTGTGTCTGC	TCTCTTTAAT	CTCTGTGAGA	GTGCCATCTC	1064
TTCTGTGAGA	CTGTGTGAGA	TGCCACTGTC	TCCTCTTTCT	GCTAACATTO	CTGTGTCTCT	1124
CTTTTGTGTA	TGTGAAATGA	TCITTTAAAO	ACT GCA GAT	AGT GGT GAA	GGT GAC	1177
Thr Ala Asp Ser Gly Gln Gly Asp 20 25						
TTT CTA GCT	GAA GGA GGA GGC	GTG COT GGC	CCA AGG GTT	GTG GAA AGA		1225
Phe Leu Ala Gln	Gly Gly Val Arg Gly	Pro Arg Val Val Gln Arg				
30 35 40						
CAT CAA TCT	GCC TGC AAA GAT	TCA GAC TGG	CCC TTC TGC	TCT GAT GAA		1273
His Gln Ser	Ala Cys Lys Asp	Ser Asp Trp	Phe Cys Ser	Asp Gln		
45 50 55						
GAC TGG	GTAAAGCATC	AGCGGGGAAA	GCAAGGAGAT	CTTTCCCTCT	GATGCTAGAG	1329
Asp Trp 60						
GGGCTCACAO	GCTGACCTGA	TTGTTCCCAO	AAACTTTTTT	AAATAAAAAA	TAATTGAATA	1389
GTTCACCTACA	TAGCAAAATA	AGAAAAAGAA	CCTACTCCCA	AGAGCACTOT	TTATTTACCT	1449
CCCAACTCTG	GGATCATTAO	TGGGTGAACA	GACAGGATTT	CAOTTGCTAT	CTCAAGCAAA	1509
ACCAAGCTCC	TGAATATTGT	GGCTCAATT	TCCTGGCACC	TATTTATGOC	TAAOTGGACC	1569
CTCATTCGAO	AGTTTCTCTG	CGACCTCTAA	CTAGTCTCTT	TACCTACTTT	TAAGCCAACT	1629
TATCTGGAAG	AGAAAAAGGTA	GGAAAAAATO	GGGCTGOCAT	GGAAACATOC	AAAAATTATC	1689
TGAATCTGAG	AGATAGATCC	TTACTGTAAT	TTTTCTCCCT	CACTTTCAO	AAC TAC	1744
Asp Trp 75						
AAA TGC CCT	TCT GGC TGC	AGG ATO AAA	GGG TTT ATT	GAT GAA	GTG AAT	1792
Lys Cys Pro	Ser Gly Cys Arg	Met Lys Gly	Leu Lys Asa	Ser Leu Val	Ala Asn	
65 70 75						
CAA GAT TTT	ACA AAC AGA	ATA AAT AAG	CTC AAA AAT	TCA CTA	TTT GAA	1840
Gln Asp Phe	Thr Asn Arg	Leu Lys Leu	Lys Asa Ser	Leu Phe Gln		
80 85 90						
TAT CAO AAO	AAC AAT AAG	GAT TCT CAT	TGO TTO ACC	ACT AAT	ATA ATO	1888
Tyr Gln Lys	Asn Lys Asp	Ser His Ser	Leu Thr Thr	Asn Ile Met		
95 100 105 110						
GAA ATT TTO	AGA GGC GAT	TTT TCC TCA	OCC AAT	A GTAAATATTA		1932
Gln Ile Leu	Arg Gly Asp	Phe Ser Ser	Ala Asn			
115 120						
CATATTACT	TCTTTOACTT	TATAACAAGAA	ACAACAAAAA	TCCTAAATAA	ATATGATATC	1992

-continued-

CCCTTATATC	TATGACAATT	TCATCCCAAA	GTACTTAOTG	TAGAAACACA	TACCTTCATA	2032
ATATCCCTGA	AAATTTTAAO	AGGGAAGCTT	TGTTTTCTT	ATTTTTTCAA	AOTAAAAOAT	2112
GTAACTGAG	ATGTTTTAAO	GTCCAAAAAT	AAOTCAOAA	TTTGGAATTAA	AACAAGAAAT	2172
TAAATGTOTT	CTTTTCAACA	GTATATACTG	AAAGTAAGAT	GGGTCAOACT	CTTTGAOTTG	2232
ATATTTTTOT	TTCTGCTTTO	TAAAGGTGAA	AACTGAGAGG	TCAAAGAACT	TGTTCAAAGA	2292
CACAGAGCTG	GAATTTCAAC	TCCCAGAGCTC	CACCTGAGCTG	ATTAGGTAGA	TTTTTAAAT	2352
TAAAAATAG	GGTCAAGCTA	CCTCATTTCTC	ACAOTCTACT	CATTAGGGTT	AGGAAACATT	2412
GCATTCACTC	TGGGCAAGGA	CAAGGAGTCT	AGGGAOTCTT	CAOTTTCTCA	AGTTTTGCTT	2472
TGCTTTTTTA	CACCTTCACA	AACACTTGAC	ATTTAAAAATC	AGTGATGCCA	ACACTAGCTG	2532
GCAAGTGAOT	GATCTGTTTO	ACCCAAAAACA	GCTTAGGAAC	CATTTCAAAT	CTATAGAGTT	2592
AAAAAGAAAA	GCTCATCAOT	AAGAAAAATCC	AATATGTTCA	AGTCCCTTGA	TTAAGGATOT	2652
TATAAAATAA	TTGAAATGCA	ATCAAAACCAA	CTATTTTAA	TCCAAATTAC	ACCTTTAAAA	2712
TTCCAAAGAA	AGTTCTTTCT	CTATATTTCT	TTGGGATTAC	TAATTTGCTAT	TAGGACATCT	2772
TAAGTGCAT	TCATGGAAGG	CTGCAAGGCA	TAACATTATC	CAAAAGCTAA	ATGCCCCATA	2832
GGTTTTGAAC	TCACAGATTA	AACTGTAACC	AAAAATAAAT	TAGGCATATT	TACAAGCTAG	2892
TTTCTTTCTT	TCTTTTTTCT	CTTTCTTTCT	TTCTTTCTTT	CTTTCTTTCT	TTCTTTCTTT	2952
CTTTCTTTCT	TTCTCTTTCC	TTCTTTCTTT	CTTTCTTTTT	TTGCTGCGAA	TTACAGACAA	3012
ATCACTCAGC	AGCTACTTCA	ATAACCATAT	TTTCAATTTC	AG	AC	3065
				Asp	Arg	123
ACC	TAC	AAC	CGA	GTG	TCA	3113
Thr	Tyr	Asp	Arg	Val	Ser	130
AGA	AGC	AAA	GGT	ATA	GAA	3161
Lys	Arg	Lys	Val	Ile	Gln	145
AAT	GTT	AGA	GCT	CAO	TTG	3210
Asp	Val	Arg	Ala	Gln	Leu	160
GGCTGTGCTC	CCGAGTGTCC	TTGTTTTTGA	GTAGAGGGGA	AAAGGAGGGC	ATAATTATGC	3270
ACTGAGTGTG	TACTATATGC	AGAGAAAAAT	GTATATATCCA	TCATCTACCT	AAAAATAGGT	3330
ATTATTTTCT	TCACCTCCACA	GTGGAAGAAA	AAAAAATTCA	GAAGATATTAA	GTAAATTTTC	3390
CAACGTACAT	AGATAATAAT	TCAAAAGCAAT	GTTCAGTCCC	TGCTATTATCC	AAAGCCATTAC	3450
ATCACCACAC	CTCTGAGCCC	TCAGCCTGAG	TTCAACCAAGG	ATCATTTTAA	TAGGCTTTCC	3510
TTTGAGAGGG	AATAGCACCT	TACTCTTGAT	CCATTCTGAG	GCTAAGATGA	ATTAACACAG	3570
ATCCATTGCT	TATCTGTGCT	AGCCCTGCAA	TACCCAAACAT	CTCTTCCACT	GAAGGTTGCT	3630
GATAGAGGCA	AAACAGAGAA	TATTAAGTGG	TAGGCTCTCC	AGTCAAAAAA	AATGAAACCA	3690
GTITTCAGAA	GGAAAAATTAA	CTACCAAGAA	CTCAATAGAC	GTAGTTTATG	TATTTGTATC	3750
TACATTTTCT	CTTTATTTTT	CTCCCTCTCT	TCTAG	GTG	GAC	3803
				Val	Asp	175
ATC	CGA	TCT	TGT	CGA	GGG	3851
Ile	Arg	Ser	Cys	Arg	Gly	180
GAT	CTG	AAO	GAC	TAT	GAA	3899

-continued

Asp	Leu	Lys	Asp	Tyr	Gln	Asp	Gln	Lys	Gln	Leu	Gln	Val	Ile		
195							200			205					
GCC	AAA	GAC	TTA	CTT	CCC	TCT	AGA	GAT	AGG	CAA	CAC	TTA	CCA	CTG	ATA
Ala	Lys	Asp	Leu	Leu	Pro	Ser	Arg	Asp	Arg	Gln	His	Leu	Pro	Leu	Ile
210						215					220				3947
AAA	Lys	Mat	Lys	Pro	Val	Pro	Asp	Leu	Val	Pro	Gly	Asa	Phe	Lys	Val
225						230					235				3995
CTT	CAQ	AAG	GTA	CCC	CAQ	GAG	TGG	AAG	GCA	TTA	ACA	AGC	ATG	CCG	CAG
Leu	Gln	Lys	Val	Pro	Pro	Gln	Trp	Lys	Ala	Leu	Thr	Asp	Met	Pro	Gln
				245					250					255	4043
ATG	AGA	ATG	GAG	TTA	GAG	AGA	CCT	GTT	GGA	AAT	GAG	ATT	ACT	CGA	GGA
Met	Arg	Met	Gln	Leu	Gln	Arg	Pro	Gly	Gly	Asa	Gln	Ile	Thr	Arg	Gly
				260					265					270	4091
GGC	TCC	ACC	TCT	TAT	GGA	ACC	GGA	TCA	GAG	ACG	GAA	AQC	CCC	AGG	AAC
Gly	Ser	Thr	Ser	Tyr	Gly	Thr	Gly	Ser	Gln	Thr	Gln	Asc	Pro	Pro	Asa
				275					280					285	4139
CCT	AGC	AOT	GCT	GGA	AGC	TGG	AAC	TCT	GCG	AGC	TCT	GGA	CGT	GGA	AGT
Pro	Ser	Sar	Ala	Gly	Sar	Trp	Asa	Ser	Gly	Ser	Sar	Gly	Pro	Gly	Sar
				290										300	4187
ACT	GGA	AAC	CGA	AAC	CCT	GCG	AGC	TCT	GCG	ACT	GGA	GCG	ACT	GCA	ACC
Thr	Gly	Asa	Arg	Asa	Pro	Gly	Sar	Sar	Gly	Thr	Gly	Gly	Thr	Ala	Thr
303						310					315				320
TGG	AAA	CCT	GCG	AGC	TCT	GGA	CCT	GGA	AGT	GCT	GGA	AGC	TGG	AAC	TCT
Trp	Lys	Pro	Gly	Ser	Ser	Gln	Pro	Gly	Ser	Ala	Gly	Asc	Trp	Asa	Sar
						325								335	4283
GCG	AGC	TCT	GGA	ACT	GGA	AGT	ACT	GGA	AAC	CAA	AAC	CCT	GCG	AGC	CCT
Gly	Ser	Ser	Gly	Thr	Gly	Sar	Thr	Gly	Asa	Gln	Asa	Pro	Gly	Sar	Pro
						340								350	4331
AGA	CCT	GCT	AOT	ACC	GGA	ACC	TGG	AAT	CCT	GCG	AGC	TCT	GAA	COC	GGA
Arg	Pro	Gly	Ser	Thr	Gly	Thr	Trp	Asa	Gly	Ser	Sar	Gln	Arg	Gly	
							360							365	4379
AGT	GCT	GCG	CAC	TGG	ACC	TCT	GAG	AGC	TCT	GTA	TCT	GCT	AOT	ACT	GGA
Ser	Ala	Gly	His	Trp	Thr	Sar	Gln	Ser	Ser	Val	Ser	Gly	Ser	Thr	Gly
							375							380	4427
CAA	TGG	CAC	TCT	GAA	TCT	GGA	AOT	TTT	AGG	CCA	GAT	AGC	CCA	GAC	TCT
Gln	Trp	His	Sar	Gln	Ser	Gly	Ser	Phe	Arg	Pro	Asp	Sar	Pro	Gly	Sar
385											395				400
GCG	AAC	GCG	AGG	CCT	AAC	AAC	CCA	GAC	TGG	GCG	ACA	TTT	GAA	GAG	GTC
Gly	Asa	Ala	Arg	Pro	Asa	Asa	Pro	Asp	Trp	Gly	Thr	Phe	Gln	Gln	Val
														415	4523
TCA	GGA	AAT	GTA	AOT	CCA	GCG	ACA	AGG	AGA	GAG	TAC	CAC	ACA	GAA	AAA
Ser	Gly	Asa	Val	Ser	Pro	Gly	Thr	Arg	Arg	Gln	Tyr	His	Thr	Gln	Lys
														430	4571
CTG	GTC	ACT	TCT	AAA	GGA	GAT	AAA	GAG	CTC	AGG	ACT	GGT	AAA	GAG	AGG
Leu	Val	Thr	Ser	Lys	Gly	Asp	Lys	Gln	Leu	Arg	Thr	Gly	Lys	Gln	Lys
														445	4619
GTC	ACC	TCT	GGT	AGC	ACA	ACC	ACC	ACG	COT	COT	TCA	TGC	TCT	AAA	ACC
Val	Thr	Sar	Gly	Ser	Thr	Thr	Thr	Thr	Arg	Arg	Sar	Cys	Sar	Lys	Thr
														460	4667
GTT	ACT	AAG	ACT	GTT	ATT	GCT	CCT	GAT	GGT	CAC	AAA	GAA	GTT	ACC	AAA
Val	Thr	Lys	Thr	Val	Ile	Gly	Pro	Asp	Gly	His	Lys	Gln	Val	Thr	Lys
														480	4715
GAA	GTG	GTG	ACC	TCC	GAA	GAT	GGT	TCT	AGC	TGT	CCC	GAG	GCA	ATG	GAT
Gln	Val	Val	Thr	Ser	Gln	Asp	Gly	Ser	Asp	Cys	Pro	Gln	Ala	Met	Arg
														495	4763
TTA	GGC	ACA	TTG	TCT	GGC	ATA	GGT	ACT	CTG	GAT	GCG	TTG	COC	CAT	AGG
Leu	Gly	Thr	Ser	Gly	Ile	Gly	Thr	Lys	Leu	Asp	Gly	Phe	Arg	His	Arg
														510	4811
CAC	CCT	GAT	GAA	GCT	GCC	TTC	TTC	AGC	ACT	GCC	TCA	ACT	GGA	AAA	ACA
															4853

-continued

His	Pro	Asp	Glu	Ala	Ala	Phe	Phe	Asp	Thr	Ala	Ser	Thr	Gly	Lys	Thr	
515						520						525				
TTC	CCA	GGT	TTC	TTC	TCA	CCT	ATG	TTA	GGA	GAG	TTT	GTG	AGT	GAG	ACT	4907
Phe	Pro	Gly	Phe	Phe	Ser	Pro	Met	Leu	Gly	Glu	Phe	Val	Ser	Glu	Thr	
530						535					540					
GAG	TCT	AGG	GGC	TCA	GAA	TCT	GCC	ATC	TTC	ACA	AAT	ACA	AAG	GAA	TCC	4955
Glu	Ser	Arg	Gly	Ser	Glu	Ser	Gly	Ile	Phe	Thr	Ala	Thr	Lys	Glu	Ser	
545					550				555					560		
AGT	TCT	CAT	CAC	CCT	GCG	ATA	GCT	GAA	TTC	CCT	TCC	CCT	GCT	AAA	TCT	5003
Ser	Ser	His	His	Pro	Gly	Ile	Ala	Glu	Phe	Pro	Ser	Arg	Gly	Lys	Ser	
				565					570					575		
TCA	AGT	TAC	AAC	AAA	CAA	TTT	ACT	AGT	AAC	ACG	AGT	TAC	AAC	AGA	GGA	5051
Ser	Ser	Tyr	Ser	Lys	Glu	Phe	Thr	Ser	Ser	Thr	Ser	Tyr	Asp	Arg	Gly	
				580				585					590			
GAC	TCC	ACA	TTT	GAA	AAC	AAG	AAC	TAT	AAA	ATG	GCA	GAT	GAG	GCC	GGA	5099
Asp	Ser	Thr	Phe	Glu	Ser	Lys	Ser	Tyr	Lys	Met	Ala	Asp	Glu	Ala	Gly	
				595			600					605				
AGT	GAA	GCC	GAT	CAT	GAA	GGA	ACA	CAT	AAC	ACC	AAG	AGA	GAC	CAT	GCT	5147
Ser	Glu	Ala	Asp	His	Glu	Gly	Thr	His	Ser	Thr	Lys	Arg	Gly	His	Ala	
	610				615						620					
AAA	TCT	CGC	CCT	GTG	AGA	GGT	ATC	CAC	ACT	TCT	CCT	TTG	GCG	AAG	CCT	5195
Lys	Ser	Arg	Pro	Val	Arg	Gly	Ile	His	Thr	Ser	Pro	Leu	Gly	Lys	Pro	
625				630					635					640		
TCC	CTG	TCC	CCC	TAGACTAAGT	TAAATATTTC	TGCACAGTGT	TCCCATGCCC									5247
Ser	Leu	Ser	Pro		645											
CCTTGACATT	CCTTCTTAAC	TCTCTGTTAC	ACGTGATTGA	AACACACTT	TTTTGTCTG											5307
TTTTTGTGCT	AGACTGTAAG	TTCCCTTGGGG	GCAGGGCCCTT	TGTCTGTCTC	ATCTCTGTAT											5367
TCCCAAATGC	CTAACAGTAC	AGAGCCATGA	CTCAATAAAT	ACATGTTAAA	TGGATGAATG											5427
AATTCCCTCTG	AAACTCTATT	TGAAGCTTAT	TAGTCAAAT	CTTTCACAT	TCAAAGTGTG											5487
TGCTATTAGA	ATTGTCACCC	AACGATTAA	TCACATTTTT	AGTATGTGTC	TCAAGTTGACA											5547
TTTAGGTCAG	GCTAAATACA	AGTTGTGTTA	GTATTAAAGT	AGCTTAAGTA	CTGTACTG											5607
TTACTTGCTA	TTAGTTTGTG	CAAATAAAT	TCCAAATACA	TTTAGGAAAA	ATCCCTTTG											5667
CAATTTGTAG	GTATAAATAA	CCGCTTATTT	GCATAAGTTC	TATCCCACTG	TAAATGCATC											5727
CTTTCCTAT	GGAGGGGAGG	AAAGGAGGAA	GAAAGAGAGG	AAAGGAGAGG	AACAGTATTT											5787
GCCTTATTTA	ATCTGAGCCG	TGCTATCTTT	TGTAAGGTTA	AATGAGAAAT	ACTTCTTCCA											5847
ACCAAGCTTAA	TTTTTTTTTT	AGACTGTGAT	GATGTCTCTC	AAACACATCC	TTCAGGTACC											5907
CAAAGTGCCA	TTTTCAATAT	CAAAGTATCC	GGATCC													5943

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 644 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met  Phe  Ser  Met  Arg  Ile  Val  Cys  Leu  Val  Leu  Ser  Val  Val  Gly  Thr
 1          5          10          15
Ala  Trp  Thr  Ala  Asp  Ser  Gly  Glu  Gly  Asp  Phe  Leu  Ala  Glu  Gly  Gly
          20          25          30
Gly  Val  Arg  Gly  Pro  Arg  Val  Val  Glu  Arg  His  Glu  Ser  Ala  Cys  Lys
          35          40          45

```

-continued

Asp	Ser	Asp	Trp	Pro	Phe	Cys	Ser	Asp	Glu	Asp	Trp	Asa	Tyr	Lys	Cys	50	55	60
Pro	Ser	Gly	Cys	Arg	Met	Lys	Gly	Leu	Ile	Asp	Glu	Val	Asa	Glu	Asp	65	70	75
Phe	Thr	Asa	Arg	Ile	Asa	Lys	Leu	Lys	Asa	Ser	Leu	Phe	Glu	Tyr	Gln	85	90	95
Lys	Asa	Asa	Lys	Asp	Ser	His	Ser	Leu	Thr	Thr	Asa	Ile	Met	Glu	Ile	100	105	110
Leu	Arg	Gly	Asp	Phe	Ser	Ser	Ala	Asa	Asa	Arg	Asp	Asa	Thr	Tyr	Asa	115	120	125
Arg	Val	Ser	Glu	Asp	Leu	Arg	Ser	Arg	Ile	Glu	Val	Leu	Lys	Arg	Lys	130	135	140
Val	Ile	Glu	Lys	Val	Glu	His	Ile	Glu	Leu	Leu	Glu	Lys	Asa	Val	Arg	145	150	155
Ala	Glu	Leu	Val	Asp	Met	Lys	Arg	Leu	Glu	Val	Asp	Ile	Asp	Ile	Lys	165	170	175
Ile	Arg	Ser	Cys	Arg	Gly	Ser	Cys	Ser	Arg	Ala	Leu	Ala	Arg	Glu	Val	180	185	190
Asp	Leu	Lys	Asp	Tyr	Glu	Asp	Glu	Glu	Lys	Glu	Leu	Glu	Glu	Val	Ile	195	200	205
Ala	Lys	Asp	Leu	Leu	Pro	Ser	Arg	Asp	Arg	Glu	His	Leu	Pro	Leu	Ile	210	215	220
Lys	Met	Lys	Pro	Val	Pro	Asp	Leu	Val	Pro	Gly	Asa	Phe	Lys	Ser	Glu	225	230	235
Leu	Glu	Lys	Val	Pro	Pro	Glu	Trp	Lys	Ala	Leu	Thr	Asp	Met	Pro	Glu	245	250	255
Met	Arg	Met	Glu	Leu	Glu	Arg	Pro	Gly	Gly	Asa	Glu	Ile	Thr	Arg	Gly	260	265	270
Gly	Ser	Thr	Ser	Tyr	Gly	Thr	Gly	Ser	Glu	Thr	Glu	Ser	Pro	Arg	Asa	275	280	285
Pro	Ser	Ser	Ala	Gly	Ser	Trp	Asa	Ser	Gly	Ser	Ser	Gly	Pro	Gly	Ser	290	295	300
Thr	Gly	Asa	Arg	Asa	Pro	Gly	Ser	Ser	Gly	Thr	Gly	Gly	Thr	Ala	Thr	305	310	315
Trp	Lys	Pro	Gly	Ser	Ser	Gly	Pro	Gly	Ser	Ala	Gly	Ser	Trp	Asa	Ser	325	330	335
Gly	Ser	Ser	Gly	Thr	Gly	Ser	Thr	Gly	Asa	Glu	Asa	Pro	Gly	Ser	Pro	340	345	350
Arg	Pro	Gly	Ser	Thr	Gly	Thr	Trp	Asa	Pro	Gly	Ser	Ser	Glu	Arg	Gly	355	360	365
Ser	Ala	Gly	His	Trp	Thr	Ser	Glu	Ser	Ser	Val	Ser	Gly	Ser	Thr	Gly	370	375	380
Glu	Trp	His	Ser	Glu	Ser	Gly	Ser	Phe	Arg	Pro	Asp	Ser	Pro	Gly	Ser	385	390	395
Gly	Asa	Ala	Arg	Pro	Asa	Asa	Pro	Asp	Trp	Gly	Thr	Phe	Glu	Glu	Val	405	410	415
Ser	Gly	Asa	Val	Ser	Pro	Gly	Thr	Arg	Arg	Glu	Tyr	His	Thr	Glu	Lys	420	425	430
Leu	Val	Thr	Ser	Lys	Gly	Asp	Lys	Glu	Leu	Arg	Thr	Gly	Lys	Glu	Lys	435	440	445
Val	Thr	Ser	Gly	Ser	Thr	Thr	Thr	Thr	Arg	Arg	Ser	Cys	Ser	Lys	Thr	450	455	460
Val	Thr	Lys	Thr	Val	Ile	Gly	Pro	Asp	Gly	His	Lys	Glu	Val	Thr	Lys	465	470	475

-continued

Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala Met Asp
 485 490 495
 Leu Gly Thr Leu Ser Gly Ile Gly Thr Leu Asp Gly Phe Arg His Arg
 500 505 510
 His Pro Asp Glu Ala Ala Phe Phe Asp Thr Ala Ser Thr Gly Lys Thr
 515 520 525
 Phe Pro Gly Phe Phe Ser Pro Met Leu Gly Glu Phe Val Ser Glu Thr
 530 535 540
 Glu Ser Arg Gly Ser Glu Ser Gly Ile Phe Thr Thr Lys Glu Ser
 545 550 555 560
 Ser Ser His His Pro Gly Ile Ala Glu Phe Pro Ser Arg Gly Lys Ser
 565 570 575
 Ser Ser Tyr Ser Lys Glu Phe Thr Ser Ser Thr Ser Tyr Asn Arg Gly
 580 585 590
 Asp Ser Thr Phe Glu Ser Lys Ser Tyr Lys Met Ala Asp Glu Ala Gly
 595 600 605
 Ser Glu Ala Asp His Glu Gly Thr His Ser Lys Arg Gly His Ala
 610 615 620
 Lys Ser Arg Pro Val Arg Gly Ile His Thr Ser Pro Leu Gly Lys Pro
 625 630 635 640
 Ser Leu Ser Pro

(2) INFORMATION FOR SEQ ID MOD:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8878 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: human Rhinogen B-base chain

(ix) FEATURE:

- (A) NAME/KEY: misc_RNA
- (B) LOCATION: 1..469

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 470..583

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 584..727

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 728..849

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 850..998

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 999..412

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 413..504

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 505..570

(ix) FEATURE:

-continued

```

      (A) NAME/KEY: intron
      (B) LOCATION: 5271_5830

      (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 5831_5944

      (ix) FEATURE:
        (A) NAME/KEY: intron
        (B) LOCATION: 5945_6632

      (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 6633_6758

      (ix) FEATURE:
        (A) NAME/KEY: intron
        (B) LOCATION: 6759_6966

      (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 6967_7252

      (ix) FEATURE:
        (A) NAME/KEY: intron
        (B) LOCATION: 7253_7870

      (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 7871_8102

      (ix) FEATURE:
        (A) NAME/KEY: 3'UTR
        (B) LOCATION: 8103_8537

      (ix) FEATURE:
        (A) NAME/KEY: misc_RNA
        (B) LOCATION: 8538_8878

      (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: join(470_583, 3258_3449, 3939_4122,
          5040_5270, 5831_5944, 6633_6758, 6967_7252,
          7871_8102)

      (xi) SEQUENCE DESCRIPTION: SDQ ID NO3:

      GAATTCATGC CCCTTTTGA ATAGACTTAT GTCATTGTCA GAAACATAAA GCATTTATGG      60
      TATATCATTAT ATGATGCACG ATTTTAAGTG TTCCCTTGTG AGTAGGTCAA ATTTACTAAG      120
      CTTAAGTTTG TTTTCTCACA TATTCTTTGC GAGCTTGTGT AGTTCCACA TTAATTTACC      180
      AGAAACAAAG TACACACTCT CTTTGAAGAG TCCCTAACT TCCATCATT TTGTCCAATT      240
      AAATGAATTO AAGAAATTTA ATGTTTCTAA ACTAGACCAA CAAAGAAATA TAGTTGTATG      300
      ACAAGTAAAT AAGCTTTGCT GGGAAAGATG TCTTTAAATG ATAAATGATG TCAAGCAACA      360
      AGTGAACCAA AAATTAATAA TTAAGTAAGG AAGAGTAACC ATTTCTGAAG TCATTCTAG      420
      CAGAGGACTC AGATATATAT AGGATTTAAG ATCTCTCAOT TAACTCTAC ATG AAA          475

                                         Met Lys
                                         1

      AGG ATG GTT TCT TGG AGC TTC CAC AAA CTT AAA ACC ATG AAA CAT CTA      523
      ATG Met Val Ser Trp Ser Phe His Lys Leu Lys Thr Met Lys His Leu
                                         5          10          15

      TTA TTA CTA CTA TTA TGT GTT TTT CTA GTT AAG TCC CAA GGT GTC AAC      571
      Leu Leu Leu Leu Cys Val Phe Leu Val Lys Ser Gln Gln Val Asn
                                         20          25          30

      GAC AAT GAG GAG GTGAATTTT TAAAGCATTAT TTATATTATT AGTAGTATTA      623
      Asp Asn Gln Gln
                                         35

      TTAATATAAG ATGTAACATA ATCATATTAT GTGCTTATTT TAATGAAATT AGCATTGCTT      683

```


ATAGTTATGA	AATGGAATTT	TTAACCTCTG	ACTTATTGTA	TTTAAAGAA	GTTCATAAT	743
ATTTCTTATA	TAAAAACAAA	GTAAATTTCT	GTTTTCTAGT	TTATCACCTT	TGTTTTCTTA	803
AGATGAGGAT	GGCTTAGCTA	ATGTAAAGAT	TGTTTTTCTC	ACTTGTCTAT	CTGAGTACTG	863
TGATTTTCAT	TTACTTCTAG	CAATACAGGA	TTACAATTAA	GAGGACAAGA	TCTGAAAAATC	923
TCACAAAACTA	TAAAAATAATA	AAAGAGCAGA	ATTTTAAAGAT	AAAAAGAACT	GGTGGTAGGT	983
AGATTGTTCT	TTGGTGAAGG	AAAGTAATAT	ATATTGTTAC	TGAGATTACT	ATTTATAAAA	1043
ATTATAACCTA	AGCTTAAAAA	CAAAATACAT	CAAGTGTAA	GATAGAAAAA	GAATATTGTC	1103
TTTTTTCAGA	TGAAAAGTTC	AAATTAGAAT	TAGTGTGTAT	TGTTATTATT	AATAGTTATG	1163
AAACACGGTT	CACTCTAATT	TATTTATTTG	TAGAACAGTT	TGTCCTCAAC	TATTTATTTT	1223
GCTGACTTAT	TGCTGTAAAT	TTGCACTTAC	TAAAAATACA	GAATGTATAT	TAGGACAATG	1283
GATATTTAAG	AAATTTAAAT	TTTATCATCA	AACGTATCAT	GGCCAAATTT	CTTACATATA	1343
GCATAGTATC	ATTAACTAG	AAATAAGAA	ACACAATAAT	ATTTAAATGA	AGTGAATTCAT	1403
TTGCGATCAT	TATGAGTTT	CAAGGGAAC	TGAGTGTGTT	ACTTATCAGA	CTCTACATGT	1463
AAAGACATAT	AGTTAACTCT	GTGTTGTGTT	TAAAAACATA	TGTTTAACTC	GGTTAAAGTCT	1523
GGTTAATCAT	ATTAGTAAAG	AAAAATGTAA	AGAAATGTGA	AGACGAATTT	TTTGTAAAGT	1583
ACTCTGCAAA	GCACCTTAC	ATTTCTGCTT	ATCAACTAAA	CTCAGCAGAG	ATAGTTTAAAT	1643
AGTTTAAAGCT	TTAAAAATGA	TTTTGATTAT	TCAACAAGTG	GGCTTCATAA	TTTCTTTAAG	1703
TGTTTTTCTT	TAAATATATA	CTTCTTTTAA	ATATTTTTTA	AAATTTCTCT	TTCTCTAGTA	1763
AAAGCAGACC	ATCCATGCTA	CTCTCTAAT	GGCACTCTGA	AATAAAAAAG	AAATAGTTTT	1823
CTCTGTTATA	ATTGTATTTG	TAAATAAGCA	ATGAATCACA	TTTCTTAAAA	TTTGTTTTTA	1883
AGAGGGTAAAG	CTCTGACTAG	GACCATGACT	TCAATGTGAA	ATATGTATAT	ATCTCCGAA	1943
TCCTTACATA	TTAAGAAAT	ATATAATCAA	CTGGTTAAAC	AGGAAAAATCT	GGAAACAACCT	2003
GGCTGGGTTT	TAACTCTAGC	ACCATCTAC	TAAATGTTAA	ATAATATTAT	AATCTAATGA	2063
ATAAATGACA	ATGCAATTC	AAATAGAGTT	CATCTGATGA	CTTCTAGACT	CACAAAAATG	2123
CAAGAGAGCT	CACTTGTGTC	TCACTTGTTC	CAAAATCATGT	CGTTTGTTTA	TTTGTAAATTA	2183
AGCTCCAAA	GATGTATAGC	TACTGACAAA	AAAAAAAATG	AGAAATGAT	TAACTCCAAAT	2243
CAAAAATCTT	CTATTGCAAT	GGCTATTTTC	TGCTTCATTA	TGCTTTAAATA	TAAATTTTTA	2303
AGTTAGCAAG	TAATTTTAAT	TACAATGCAC	AAAGCTTGAAG	AATTTATTTA	AATATAAGAA	2363
AATCATAATG	TTTGATAAAG	AAATCATGTA	AGAAATTTCA	AGATAATGAT	TTAACAAATA	2423
ATTTTGTGTA	TAGAAAGATA	GACTAAAAAT	GAATTTGCAA	GTGGAGAGGA	CACTTAAACT	2483
GTAGTACTTG	TTATGTGTGA	TTCCAATGAA	AATAGTAATG	AGCACTTATT	ATTGCCAAGT	2543
ACTGTCTGGA	GGGTACCATA	TGCAATAAGT	TATTTAATCC	TTACAATTAAT	CTTGTAAGGC	2603
AGATTCAACAC	TATCATATCA	CTTATTTTAC	AGATGAGAAA	ACTGGGGCAC	AGATAAAGCA	2663
ACTTGCCCAA	GGTCTCATAG	CTGTAAATCA	ACCCCTACGGT	CAAGACCTAC	AAATAGCCGA	2723
GCTCCAGAGT	ACATTATGAG	GGTCAAAAGT	TGTCTTATTA	CAAAATAAAT	CCAAAGTAGAA	2783
TCAACCTTTA	ATAAGTCTTT	AATGTCTCTT	AAATATGTTT	ATATAAGAGT	CTAATCACCA	2843
ATTCACAAAA	ATGAAAGTGA	GGAAATGATT	AACAATAATC	ATAAGAAATC	AACAATCCAA	2903
GTGGCTTGAG	AAATATTCATT	CTTCTTGACA	GTATAGATTC	TTTACAATTT	CGTAAAGTTCC	2963
AATGTATGTT	TTAGGAATAT	GAGGTCTATTA	CTATTTCATA	TCTGATACAG	CTTTATCCCTA	3023
AGGCCCTCTCT	TTAAAAACTA	CACGTGCATCA	TAGCTTTTTT	GTGCAAGTTGG	TCTTTCTACT	3083

OTTACTGAAAC AGTAAAGCAAC CTACAGATTG ACTATACACCA ACCAAGCAATG TGATGGATCT	3143
TAAGCAAATT ATCAAGCTTG TGATAACCTA AATTATATAAAA TGAGGGTOTT GGAATAAGTTA	3203
CATTCCAAAT CTCTATAAC ACTCTGTATT ATATTCTGCG CTCATTCTCT GTAG GGT	3260
Oly	
TTC TTC AGT GCC CGT GGT CAT CGA CCC CTT GAC AAG AAG AGA GAA GAG	3308
Phe Phe Ser Ala Arg Gly His Arg Pro Leu Asp Lys Lys Arg Glu Glu	
40 43 50 53	
GCT CCC AGC CTG AAG CCT GCC CCA CCG CCC ATC AGT GGA GGT GGC TAT	3356
Ala Pro Ser Leu Arg Pro Ala Pro Pro Ile Ser Gly Gly Tyr	
60 63 70	
CGG GCT CGT CCA GCC AAA GCA GCT GCC ACT CAA AAG AAA GTA GAA AGA	3404
Arg Ala Arg Pro Ala Ala Ala Thr Glu Lys Val Glu Arg	
75 80 83	
AAA GCC CCT GAT GCT GGA GGC TGT CTT CAC GCT GAC CCA GAC CTG	3449
Lys Ala Pro Asp Ala Gly Gly Cys Leu His Ala Asp Pro Asp Leu	
90 93 100	
GTGGGTGCAC TGATGTTTCT TGCAAGTGGTG GCTCTCTCAT GCAGAGAAAG CCTGTAGTCA	3509
TGGCAGTCTG CTAAATGTTT ACTGACCCAC ATTACCATCA CTGTATTATT GTTTGTTTAT	3569
TTTGGAAATA AAATTTCAAAA CATAAACATA TTGGGCCCTT GGTTAAGGCT TTCTTTCTTG	3629
TTTCTTTTGG TCTGGGCCCA AAATTTCAAA TTAGGATATG TGGGTGCCAC CTCTCCATT	3689
GTATTTTGCC ACTGCCCTTG TTTAGTTGGT AAAATTTTCA TAGCCCAATT ATATTTTTC	3749
TGGGTAAAT AATATTTTAA ATCTCTATGA GAGTATGATG ATGACTTTTC AATTTCTGGT	3809
CTTACAGAAA ACCAAATAAT AAATTTTAT GTTGCTAAT CGATCCCTG AATTTTCTTA	3869
TGTGCTATTT TAACAAATGT CCATGACCCA AATCCTTCAT CTAATGCCCT CTATTTTCTT	3929
TGTTTTAG GGG GTG TTG TGT CCT ACA GGA TGT CAG TTG CAA GAG GCT	3977
Gly Val Leu Cys Pro Thr Gly Cys Glu Ala Glu Ala	
105 110 115	
TTG CTA CAA CAG GAA AAG CCA ATC AGA AAT AGT GTT GAT GAG TTA AAT	4025
Leu Leu Glu Glu Glu Arg Pro Ile Arg Asn Ser Val Asp Glu Leu Asn	
120 123 130	
AAC AAT GTG GAA GCT GTT TCC CAG ACC TCC TCT TCT TCC TTT CAG TAC	4073
Asn Asn Val Glu Ala Val Ser Glu Thr Ser Ser Ser Ser Phe Glu Tyr	
135 140 143	
ATG TAT TTG CTG AAA GAC CTG TGG CAA AAG AAG CAG AAG CAA GTA AAA	4122
Met Tyr Leu Leu Lys Asp Leu Thr Glu Lys Arg Glu Lys Glu Val Lys	
150 153 160	
GTAGATATCC TTGTGCTTTC CATTGATTT TCACTATATA AATTGGAACC GTTAGACTGC	4182
CACGAGAATG CATGGTTGTG AGAAGATTAA CATTCTGGGG TTAGTGAATA GCATTCATAC	4242
GCTTTTGGGC ACCTTCCCTT GCAACTTGCC AGATAAGCAC TATTAGCTC TTATTTCCAG	4302
TCTGACATCA GCAAGGTGGA TTTTCTATGA AAAATTTCTAC TATGACTCCT TATTTTAAAT	4362
ATACAAGAAA CTGTGACTC AGAAGATAAT ATTTACAGAG TGGAAAAAAA CCCCTAGCAT	4422
TTATAGTTTT AACATTTGAG GTTTTGAATG AGAAGATTAT CCATAATATA TTCAATTGTG	4482
TTGTGGATAA TGACACCTAA CCTGTGAATC TTGAGGTGAG AATGTTGATG GCTGTTGACT	4542
TGTTGTTGAG GAAACAGCTA GTGCTGTAGC CTGGCACAGG CATCTCAAGT AGTAGCATAC	4602
CCACAGTTGG AAATTTTTC AAGAAATCAA AGGAATCATG ACATCTTATA AATTTCAAGG	4662
TTCTGCTATA CTATGTGAAA ATGGAATAAT AAATCAAGCA TATCCACTCT GTAAAGATTGA	4722
ACTTCTCAGA TGGAGAGCCC CAATACTGCT TTCTCTCTCT TTCCCTCACC AAAAGAAATA	4782
ACAACCTATT TCATTTATTA CTGGACACAA TCTTTAGCCT ATACCTATGG TAAATTACTA	4842

-continued

GTATGTTGTT TAGGATTAT GTTAATTGTT ATATGTCATG CCCCACAAATCA TTTCCTACTAA	4902
ATATGACTAT ATATCATAAC TGCTTGGTGA TAGCTCAAGT TTTAATAAGT TATTCTCAGA	4962
AAATCAAAAT TOTATAAGTTA AATACATTAG TTTTATGAGG CAAAAATGCT AACTATTTC	5022
ACATAATTTC ATTTTTCAG AT AAT GAA AAT GTA GTC AAT GAG TAC TCC	5071
Asp Asn Glu Asn Val Val Asn Glu Tyr Ser	
165 170	
TCA GAA CTG GAA AAG CAC CAA TTA TAT ATA GAT GAG ACT GTG AAT AGC	5119
Ser Glu Leu Glu Lys His Glu Leu Tyr Ile Asp Glu Thr Val Asn Ser	
175 180 185	
AAT ATC CCA ACT AAC CTT COT GTG CTT COT TCA ATC CTG GAA AAC CTG	5167
Asn Ile Pro Thr Thr Asn Leu Arg Val Leu Arg Ser Ile Leu Glu Asn Leu	
190 195 200 205	
AGA AGC AAA ATA CAA AAG TTA GAA TCT GAT GTC TCA GCT CAA ATG GAA	5215
Arg Ser Lys Ile Glu Lys Leu Glu Ser Asp Val Ser Ala Glu Met Glu	
210 215 220	
TAT TOT CAC ACC CCA TGC ACT GTC AGT TGC AAT ATT CCT GTG GTG TCT	5263
Tyr Cys Arg Thr Pro Cys Thr Val Ser Cys Asn Ile Pro Val Val Ser	
225 230 235	
GGC AAA G GTAAGTATT CATAACATA TTTTATGAGA GTTCCAGAGG AACTCACACA	5320
Gly Lys	
CCAAAAATAA GAGAACAAACA ACAACAACAA AAATGCTAAG TGGATTITTC CAACAGATCA	5380
TAATGACATT ACAATACATC ATAAAAATAT CCTTAAGCCAG TTGTGTTTTG GACTGGCCCTG	5440
GTGATTTTGC TGGTTTTGAT GAGCAGGATG GGGCACAAGT AATCCCAAGG GTGGCTGATG	5500
TGTGATCTG COTACTGGCT TGAACAGATG GCAAGAACAC AGATAGATGT AGAAGTTTCT	5560
CCATTTTGTG TOTTTCTGGA GTCATGGAT ATTCCAAGAC ACAAAAAGGTG GAGAAAGAGCT	5620
TTGTTCATCC TCTTAGCAGA TAAACCTCCT CAAAAGTGGG TTGGACTTAC TAAAGTAAAA	5680
TGAAAACTCA ATATTITGTTA TATTATTTTC AAAAGTCTAT AATAACCTAC CCTTAAATGA	5740
CTTATGTAAT GTTATTTTAA AGAATTGGTG ACTAAATACA AAGTAATTAT GTCATAAACC	5800
CCTGAACATA ATGTTGTCTT ACATTTCAGG AA TOT GAG GAA ATT ATC AAG AAA	5853
Glu Cys Glu Glu Ile Ile Arg Lys	
240 245	
GGA GGT GAA ACA TCT GAA ATG TAT CTC ATT CAA CCT GAC AGT TCT GTC	5901
Gly Gly Glu Thr Ser Glu Met Tyr Leu Ile Glu Pro Asp Ser Ser Val	
250 255 260	
AAA CGG TAT AGA GTA TAC TOT GAC ATG AAT ACA GAA AAT GGA G	5944
Lys Pro Tyr Arg Val Tyr Cys Asp Met Asn Thr Glu Asn Gly	
265 270 275	
GTAAAGCTTC GACAGTTGTT GACCTGTTGA TCTGTAAATTA TTTGGATACC GTAAAAATGCC	6004
AGGAAACAGG GCCAGGTGTG GTGGCTCATA CCTGTAAATT CAACACCTTG GAGAGCCAAA	6064
GTGGGCTGAT AGCTTGAGCC TAGGAGTTTG AAAGTACGCT GGGCAACATA ATGAGACCTT	6124
AACTCTACAA AAAAAAAAAA AATACCAAAA AAAAAAAAAA AATCAAGCTGT GTTGGTAGTA	6184
TGTGCTGTGA GTCCCAAGCTA TCCAGGAGGC TAGAGATGGGA GATCAGCTGA GCCCACAACC	6244
TGGAGTCTTG ATCATGCTAC TGAAGTGTAG CCTGGGCAAC AGAGGATAGT GAGATCCTGT	6304
CTCAAAAAAA AAAATTAATT AAAAAAGCCAG GAAACCAAGAC TTAAGCTCTAA CATCTAACAT	6364
AGCTGACAAA GGAATTAATT GATGTGGAA TCAACCTGAT ATTTAAAAAT TATAAAATAT	6424
CTATAATTCA CAATTGAGGG TAGAGTAAAG CACTTGCAAG TTCCAAAGAT TTTACAAATT	6484
TACCTCTCAT ATTTATTTTC TTATTGTGTC TATTTTAGAG CACCAAAATAT ATACTAAATG	6544
GAATGAGCAG GGAATTCAGA TATTATTTTC AAAAGTACAT TATTTGCTGT TGGTTAATAT	6604

-continued-

ATGCTCTTTT	TGTTTCTGTC	AACCAAAO	GA	TGG	ACA	GTG	ATT	CAO	AAC	COT	6655
			Oly	Trp	Thr	Val	Ile	Ola	Asa	Arg	
											285
CAA	GAC	GGT	AGT	GTT	GAC	TTT	GCC	AGG	AAA	TGG	6703
Ola	Asp	Gly	Ser	Val	Asp	Phe	Gly	Arg	Lys	Trp	
											290
OGA	TTT	OGA	AAT	GTT	OCA	ACC	AAA	ACA	GAT	GGG	6751
Gly	Phe	Gly	Asa	Val	Ile	Thr	Asa	Thr	Asp	Oly	
											305
CTA	CCA	G	GTAAACGAACA	GGCATGCAAA	ATAAAAATCAT	TCTATTTTGA	ATGGAATTTT				6800
Leu	Pro										
TTTTAAATTA	AAAACATTC	TTGTTGGAAG	CCTGTTTTAG	GCAGTTAAAG	OGAGTTTCCT						6868
GACAAAAAT	TGGAAGCTAA	AGATAAGGGA	AQAAAAAGCA	TTTTTAATTT	CCCAAAATTT						6928
TATTTTTGOT	GAGAGATTTT	ATTTTTTTTT	TCTTTTGA	GT	GAA	TAT	TGG	CTT			6980
OGA	AAT	GAT	AAA	ATT	AGC	CAO	CTT	ACC	AGG	ATG	7028
Gly	Asa	Asp	Lys	Ile	Ser	Ola	Leu	Thr	Arg	Met	
											320
TTT	ATA	GAA	ATG	GAG	GAC	TGG	AAA	OGA	GAC	AAA	7076
Leu	Ile	Ola	Met	Ola	Asp	Trp	Lys	Gly	Asp	Lys	
											335
OGA	OGA	TTT	ACT	GTA	CAO	AAT	GAA	GCC	AAC	AAA	7124
Gly	Gly	Phe	Thr	Val	Gla	Asa	Ola	Ala	Lys	Tyr	
											350
AAC	AAA	TAC	AGA	OGA	ACA	GCC	GGT	AAT	GCC	CTC	7172
Asa	Lys	Tyr	Arg	Gly	Thr	Ala	Gly	Asa	Ala	Leu	
											365
CAO	CTO	ATG	OGA	GAA	AAC	AGG	ACC	ATG	ACC	ATT	7220
Ola	Leu	Met	Oly	Ola	Asa	Arg	Thr	Met	Thr	Ile	
											380
TTT	AGC	ACG	TAT	GAC	AGA	GAC	AAT	GAC	GGT	TO	7262
Phe	Ser	Thr	Tyr	Asp	Arg	Asp	Asa	Asp	Oly	Trp	
											405
CACCTCTTGC	TCTGCTTTA	AAAAATCACAC	TAATATCATT	ACTCAAGATC	ATTAACAATA						7322
TTTTTAATAG	CTACCACTTC	CTGCGCACCTT	ACTGTCAGCC	ACTGTCTCTAA	GTCTTTTATG						7382
CATCACTCGA	AAGCATTTCA	ACTATAAGGT	AGACATTTCT	ATCTCTCATTT	TACAGATGAG						7442
ATTTAAGAGAG	ATTACGCTGT	TTGTCCCAAT	TCACACAACAT	ACCCCAAGATAT	AAAACATAAG						7502
TTGAAGCACA	GTTCATTTCT	GATAAATGAG	CATTATAGATA	AATACCTATA	TCTCTATATT						7562
TTGAAGGTGT	TGTGAAAACT	TGATTTTCTA	TTTCCAGGOT	TCTCTGTATAC	TAGAGGTTGT						7622
AAAAAGCTATT	ATTCCAGTAT	AAAATAACAA	ACACAGTCCC	TAGATGATTT	GCCACAAAAG						7682
CCCAATTATC	TCTCTTTCTT	GCTATAAGGC	ACAGGAGGTC	TTTGTGTAT	TAGTGTGACT						7742
CTATGTATAG	CACCCAAAAG	AAAAGACTACT	GTGCACACGA	GTGTAGCAAT	CTTTTATGGG						7802
TAATCTGCAA	AACGTAACTT	GACCACCGTA	GTTCCTTTTC	TAATAACGCC	AAACACATTT						7862
TCTTTTCAO	G	TTA	ACA	TCA	GAT	CCC	AGA	AAA	GAC	TGT	7910

-continued

445	450	455	460	
ACA GAT GAT GGT GTA GTA TGG ATG AAT TGG AAG GGG TCA TGG TAC TCA				8054
Thr Asp Asp Gly Val Val Trp Met Asa Trp Lys Gly Ser Trp Tyr Ser				
	465	470	475	
ATG AAG AAG ATG AGT ATG AAG ATC AAG CCC TTC TTC CCA CAG CAA TAGTCCCC				8109
Met Arg Lys Met Ser Met Lys Ile Arg Pro Phe Phe Pro Glu Glu				
	480	485	490	
TACGTAGATT TTGCTCTTC TGTATGTGAC AACATTTTTT TACATTATGT TATTGGAATT				8169
TTCTTTCATA CATTATATTC CTCTAAAACCT CTCAGCAGA CGTGAAGTGT ACTTTTTGAA				8229
AAAAGTATAG GATAAATTAC ATIAAAATAO CACATGATTT TCTTTTGTIT TCTTCATTTC				8289
TCTTGCTCAC CCAAAGAACTA AAAAAAGTAT AGTTTTGACA GAGTTGGTGT TCATAATTTC				8349
AGTTCTAGTT GATTGCGAGA ATTTTCAAAT AAGGAAGAGG GGTCTTTTAT CCTGTGCTGA				8409
GGAAAACCAT GACGGAAAAG AAAAAGTGT GTTTAAAAGT CCACTTTTAA AACTATATTT				8469
ATTTATGTAG GATCTGTCAA AGAAAAGTTC CAAAAAGATT TATTAATTAA ACCAGACTCT				8529
GTTGCAATAA GTTAATGTIT TCTTGTITTO TAATCCACAC ATTCAGTGAO TTAGGCTTTO				8589
CACITGTAAO GAAAGGAAAO COTTCACAAC CTCAGAAAGC TAATAAACCO GTCTTGAAAT				8649
TTTGAAGATT TAAAATCTGA CTCTAGGAGC GGCACGGTGG CTCACGACTA TAATCCCAAC				8709
ACTTTGGGAG GCTGAGGCGO GCGGTACAAA GGTCAAGAGT TCAAGACCAO CCTGACCAAT				8769
ATGTTGAAAC CCCATCTCTA CTAAAAATAC AAAAAATTAG CAGGCGTGGT GGCAGGTGCC				8829
TGTAGGTCCC AGCTAGCCTG TGAAGTGGAG ATTGCATTGA GCCAAGATC				8878

(2) INFORMATION FOR SEQ ID NO:4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 491 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4

Met	Lys	Arg	Met	Val	Ser	Trp	Ser	Phe	His	Lys	Leu	Lys	Thr	Met	Lys
1				5						10				15	
His	Leu	Leu	Leu	Leu	Leu	Leu	Cys	Val	Phe	Leu	Val	Lys	Ser	Glu	Gly
				20				25					30		
Val	Asp	Asp	Asa	Glu	Glu	Gly	Phe	Phe	Ser	Ala	Arg	Gly	His	Arg	Pro
				35			40					45			
Leu	Asp	Lys	Lys	Arg	Glu	Glu	Ala	Phe	Ser	Leu	Arg	Pro	Ala	Pro	Pro
				50			55				60				
Pro	Ile	Ser	Gly	Gly	Gly	Tyr	Arg	Ala	Arg	Pro	Ala	Lys	Ala	Ala	Ala
				65		70				75					80
Thr	Glu	Lys	Lys	Val	Glu	Arg	Lys	Ala	Phe	Asp	Ala	Gly	Gly	Cys	Leu
				85					90					95	
His	Ala	Asp	Pro	Asp	Leu	Gly	Val	Leu	Cys	Phe	Thr	Gly	Cys	Glu	Leu
				100				105					110		
Glu	Glu	Ala	Leu	Leu	Glu	Glu	Glu	Arg	Pro	Ile	Arg	Asa	Ser	Val	Asp
				115				120				125			
Glu	Leu	Asa	Asa	Asa	Val	Glu	Ala	Val	Ser	Glu	Thr	Ser	Ser	Ser	Ser
				130			135					140			
Phe	Glu	Tyr	Met	Tyr	Leu	Leu	Lys	Asp	Leu	Trp	Glu	Lys	Arg	Glu	Lys
				145			150			155					160
Glu	Val	Lys	Asp	Asa	Glu	Asa	Val	Val	Asu	Glu	Tyr	Ser	Ser	Glu	Leu
				165					170					175	

-continued

```

Glu Lys His 180  Leu Tyr Ile Asp 185  Glu Thr Val Asa Ser Asa Ile Pro
Thr Asa Leu Arg Val Leu Arg Ser 195  Leu Glu Asa Leu Arg Ser Lys
195
Ile Glu Lys 210  Leu Glu Ser Asp 215  Val Ser Ala Glu Met 220  Glu Tyr Cys Arg
210
Thr Pro Cys Thr Val Ser Cys Asa 225  Ile Pro Val Val Ser Gly Lys Glu
225
Cys Glu Glu 245  Ile Arg Lys Gly Gly Glu Thr Ser Glu Met Tyr Leu
245
Ile Glu Pro Asp Ser Ser Val Lys Pro 260  Tyr Arg Val Tyr Cys Asp Met
260
Asa Thr Glu 275  Asa Gly Gly Trp Thr Val Ile Glu Asa Arg Glu Asp Gly
275
Ser Val Asp Phe Gly Arg Lys Trp Asp Pro Tyr Lys Glu Gly Phe Gly
290
Asa Val Ala Thr Asa Thr Asp Gly Lys Asa 305  Cys Gly Leu Pro Gly
305
Glu Tyr Trp Leu Gly Asa Asp Lys Ile Ser 325  Glu Leu Thr Arg Met Gly
325
Pro Thr Glu Leu Leu Ile Glu Met 340  Glu Asp Trp Lys Gly Asp Lys Val
340
Lys Ala His 355  Tyr Gly Gly Phe Thr Val Glu Asa Glu Ala Asa Lys Tyr
355
Glu Ile Ser Val Asa Lys Tyr Arg Gly Thr Ala Gly Asa Ala Leu Met
370
Asp Gly Ala Ser Glu Leu Met Gly Glu Asa Arg Thr Met Thr Ile His
385
Asa Gly Met Phe Phe Ser Thr Tyr Asp Arg Asp Asa Asp Gly Trp Leu
405
Thr Ser Asp Pro Arg Lys Glu Cys Ser Lys Glu Asp Gly Gly Gly Trp
420
Trp Tyr Asa Arg Cys His Ala Ala Asa Pro Asa Gly Arg Tyr Tyr Trp
435
Gly Gly Glu Tyr Thr Trp Asp Met Ala Lys His Gly Thr Asp Asp Gly
450
Val Val Trp Met Asa Trp Lys Gly Ser Trp Tyr Ser Met Arg Lys Met
465
Ser Met Lys Ile Arg Pro Phe Phe Pro Glu Glu
485

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10564 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) IMMEDIATE SOURCE:

- (B) CLONE: human *Shigella flexneri* genome clone

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 398(1799-1876, 1973-2017, 2207-2290, 2510
... 2602, 4211-4241, 4665-4778, 5758-5942, 7426
... 7702, 9343-9371)

-continued-

(x1) SEQUENCE DESCRIPTION: SEQ ID NOS:

CTACACACTT	CTTGAAGCCA	AAGGCAATOC	TOAAGTCACC	TTTCATOTTC	AAATCATATT	60
AAAAAGTTAG	CAAGATOTAA	TTATCAOTOT	ACTATOTAAA	TCCTTGTAAA	TOATCAATAA	120
TTACATATTT	TCATTATATA	TATTTTAOTA	GATAATATTT	ATATACATTC	AACATTTCTAA	180
ATATAAGAAA	TTTACAAGAA	AAAAATAAAG	CTTTTTTTCC	AATCCTGTCC	TCCACCTCTG	240
CATCCCATTC	TTCTTCACAG	AAGCAACTGA	TTCAAOTCAT	TACATAOTTA	TTGAOTOTTA	300
ACTACAACATA	TOTTAAGTAC	AGCTATATAT	OTTAGATGCC	GTAGCCACAG	AAATCAOTTT	360
ACAATCTAAT	GCAOTGGATA	CAGCATOTAT	ACATATAATA	TAAGOTTOCT	ACAAATGCTA	420
TCTGAAGTAA	AAGCTTTTGA	AAAGATACTA	ATACTTAAAT	GTTTAATTCA	ACTGACTTGA	480
TTGACAACCT	ATTAGCTGAG	TGAAAAAGAT	GGATGAGAAA	GATTGTGAGA	CTTAATTGOC	540
TGOTGGTATG	GTGATATGAT	TGACAATAAC	TOCTAAOTCA	GAGAGGGGATA	TATTAAGGAG	600
GAGAGAGAAA	GCAACAAATC	TGTTTTGTAT	GTGTTCACTT	TGTTATAATT	ATGATTTATT	660
TACTGAAAT	GAATATTTAT	CTTTGTTTTT	GAOTCAA: A	ATATACCTTT	GTAAACACAG	720
AATTAAGATA	TTAGTATTTT	TTTCAAACCT	GAGGCATTTT	TCCCACTAAC	ATATTTTCATC	780
AAAACCTTATA	ATAAGCTTGG	TTCCAGAGGA	AGAAATGAGG	GATAACCAAA	AATAGAGACA	840
TTAATAATAG	TGTAAGGCC	AGTGAATAAT	CTCAATAAGC	AOTGATGACA	GACATGTTTT	900
CCCAACACACA	AGGATGCTGT	AAAGGCCAAA	CAGAAATGAT	GGCCCTCTCC	CAGCACCTCA	960
TTTTGCCCTCT	TCCTTCAOCT	ATGCCCTTAC	TCCTCTTTAG	ATACAAGGGA	GTTGATTTT	1020
TCCTCTCTCT	GAGATAGCTT	GATGGAACCA	CAGGAACAA	GAAGTGCGCT	CTTGCTCTCT	1080
TTCTGTGTGG	CAGATGGGCT	GCCATGCCCA	CTTTCAGACA	AAAGGAAAGAT	TGAGCTCAAA	1140
AGCTCCCTGA	GAGGTGAGAG	CCTATGAAAC	TGTTTGACAC	AGAGGGACAG	GAATGTATTT	1200
CCAAGGTGAT	TCATTCCTGG	GAATAGTGAA	CTGGGACATG	GGGGAAOTCA	GTCTCTCTCT	1260
GCCACAGCCA	CAGATTAATA	ATAATAATGT	TAAGTATGCC	CTAGGCTAAA	ATAATAGTGT	1320
TAAGTATGCC	CTAAGCTAAG	AAAGTTCTTT	TGATTAATCA	GGTATGGA	GACAGGACCA	1380
TCITAAAGAT	AGACTAAGTT	TOCTTAACTT	GAGGTCTATAT	CTGTTTGTCT	TCAAGCATGT	1440
ACTGGAAGAA	GTGATCATCA	ACAAGCTCCA	GGAGTCCCT	CTCTCTCACA	GCAATGGATA	1500
ATGCTTCACT	AAGCTTTTGA	GATAATTTTG	GATCAAGGAA	AAACCTTTGA	GCTGGGCCAA	1560
AAAGGAAGAG	CTTCAACCTG	TGTGAAAAAT	CTGGGAACCT	GACAGTATAG	GTGAGGGGCC	1620
AGGATGAGGA	AAAAAGGAAC	GGAAAGACCT	GGCACCCTTT	CTGTAAAGGA	GGCCCGGTGA	1680
TCAAGTCCAG	CCATTGCGAG	TCCTGGCTAT	CCCAAGAGCT	TACATAAAGG	GACAAATTGGA	1740
GCCTGAGAGG	TGACAGTGTCT	GACACTACAA	GGCTCGGAGC	TCCGGGCACT	CAGACATC	1798
ATG AGT TGG TCC TTT CAC CCC CGG AAT TTA ATT CTC TAC TTC TAT OCT						1846
Met Ser Trp Ser Leu His Pro Arg Asn Leu Ile Leu Tyr Phe Tyr Ala						
1 5 10 15						
CTT TTA TTT CTC TCT TCA ACA TOT GTA GCA GTAAATGTGCT TCCTTACAAA						1896
Leu Leu Phe Leu Ser Ser Thr Cys Val Ala						
20 25						
ACGTTGTTTA AAATGGAAAA CTGAAAAATA AAACAAGATAA TAAACTAGTG AAATTTTCGT						1956
ATTTTTTCTC TTTTAA TAT GTT OCT ACC AGA GAC AAC TGC TOC ATC TTA						2003
Tyr Val Ala Thr Arg Asp Asn Cys Cys Ile Leu						
30 35						
GAT GAA AGA TTC GTAAAGTATT TTTATGTTT TCCCTTTGTG TGTGAACTGG						2057
Asp Glu Arg Phe						

40		
AGAAGGCGCAO	AGGAATAGAA	ATAATTCCCT CATAAATATC ATCTGGCACT TOTAACITTT 2117
TAAAAACATA	GTCTAAGTTT	TACCTATTTT TCTTAATAGA TTTTAAGAT AGCATCTGT 2177
TACATTTTAT	ATCACTGTAT	TATTTTCAO GGT AGT TAT TOT CCA ACT ACC TOT 2230
		Oly Ser Tyr Cys Pro Thr Thr Cys 43
GGC ATT OCA GAT TTC CTO TCT ACT	TAT CAA ACC AAA GTA GAC AAG GAT	2278
Oly Ile Ala Asp Phe Leu Ser Thr Tyr Glu Thr Lys Val Asp Lys Asp		
50	55	60 65
CTA CAO TCT TTG GAA GAC ATC TTA	CAT CAA GTT GAA AAC AAA ACA TCA	2326
Leu Glu Ser Leu Glu Asp Ile Leu His Glu Val Glu Asa Lys Thr Ser		
70	75	80
GAA GTC AAA CAO CTO ATA AAA GCA	ATC CAA CTC ACT TAT AAT CCT GAT	2374
Oly Val Lys Glu Leu Ile Lys Ala Ile Glu Leu Thr Tyr Asa Pro Asp		
85	90	95
GAA TCA TCA AAA CCA A	GTGAGAAAAAT AAAAGACTACT GACCAAAAAA	2420
Glu Ser Ser Lys Pro		
100		
TAATAATAAT	AATCTGTGAA	GTTCITTTTC TTTTGTITTA GTTGTICTAT TTGCTTAAGG 2480
ATTTTTATGT	CTCTGATCCT	ATATTACAO AT ATO ATA GAC GCT GCT ACT TTG 2532
		Asa Met Ile Asp Ala Ala Thr Leu 103 110
AAG TCC AGG ATA ATO TTA GAA GAA	ATT ATO AAA TAT GAA GCA TCG ATT	2580
Lys Ser Arg Ile Met Leu Glu Glu Ile Met Lys Tyr Glu Ala Ser Ile		
115	120	125
TTA ACA CAT GAC TCA AGT ATT	CG GTAAAGATTT TTGTTTTAAT TTGCTCTGCA	2633
Lys Thr His Asp Ser Ser Ile Arg		
130		
AGACTGATTT	AGTTTTTATT	TAATATTCTA TACTTGAAGT AAGATTAATT TTAATGTGTT 2693
TTCCCCATTT	ATAATATCCC	AGTGACATTA TCCCTGATTA TTTTGAOCAT AGTAGAAGATA 2753
GAAOTTTTTA	GTGCAATATA	AATTATACTO GGTATATAAT GCTTATTAAT AATCACATTG 2813
AAGAAAAGAT	TTCTAGATGT	CTTCAAAATGC TAGTTTGAAC ATATTATCA AAAATTTTTT 2873
CCCCATCCCC	CATTTATCTT	ACAACATATA ATCAATCTCA TAGGAATTTG GGTOTTTGAAA 2933
ATAAAACTCT	CTTTATAAAA	ATGCTGACAA ATTGTTGTTT AAAAAATTA GCAAGCAGAG 2993
GCATAGATAA	GTTTTTGCTT	CCTAAAGTAA ATTATATTGA ATGTGGAACA GGAAGAAACA 3053
TGTCTTGAAG	GACTAAGTGT	GGCAAAATATT GCAAAOCTCA TATTGATCAT TGCAGAATGA 3113
ACCTGCATAO	TCTCTTCCCT	TCATTTGAAA GTGAATGTCT CTGTAAAGC TTCTCAAGGA 3173
CTCATAAACT	TTCTGAACAT	AAAGTCTCAO ATACAOTTTT AATATTTTTT CCCAATTTTT 3233
TTTTCTGAAT	TTTTCTCAAA	GCAAGCTTGAO AAATTTGAGAT AAATAAGTAG TAGGGAAGAG 3293
TGGCCCAAGA	AGATTTTCT	CTCTTTTTTC TATCAGAGGG CCCCTTTTAT TATTGTTATT 3353
ATTATTACTT	GCATTATTAT	TGTCCATCAT TGAAGTTTGA GGAAGTTTAT GTACAGAAAT 3413
TGCTAAGAC	AAAGTGAAGG	GAAGAACGTGG ACAAATAATT TGTCTACCCCT TTTTACTTTC 3473
AAAGAAAGAA	CGTTTATATC	ATGTGAGACA GTTTTCTATC ATTTTTGAGT ATTTGCAAGC 3533
CACCCGTATA	GTAACTACAA	AAAGAGGGTT TTTACTTCCC CCAAGTCCAT CCCAAAGCTA 3593
TOTAACCAAG	AGCATTAAG	AAAGAAAGGG AGGTATCTGT TGTTTTTATT TACATACAA 3653
AAGTTTCCAG	ATCATGTCCC	TGTGTAAATT ATATTTTGA TTGAAGCTTA TATGTATAGC 3713
CTCAATAGAT	CCACAAGTGA	AAAGTATACT CCTTCAGCAC ATGTGAATTA CTGAAGTGAO 3773

CTTTTCTCTG	TTCATAAGCA	TCAAGGCGGT	TTCCTATTAA	CCAAGTCTCG	CACTCTTCCA	3833
GTTTGTCTATC	TGCTGTCCCT	TATGCATAAA	GTAAAAAGCA	AAATGTCAAT	GACATTGTCT	3893
TATTGCACAG	GACTTTGTTA	TTTGTGTTGG	GAAGTGAAGC	AAATATGCCC	ATTCTAAGTA	3953
AAAAAGATTCA	GTCTCCACAT	GTATTCCTGT	TTTAATTGAT	TTTTTGATTT	GTITTTCTTT	4013
TTCAAAAAAT	TTATAATTTT	AAATCATGTT	AAATTAAGTA	TATAATTTTA	CATTTTCTCT	4073
AAGAATGAAA	TAATTTATCA	GAAAAGCACT	CTTAAGAAAA	TACTTAGCAG	TTCCAAAAAG	4133
AAATATAAAA	TACTCTTCT	GAAAAGAAAT	CTTATTTTTG	TCTTCTTATT	TTTGTATCT	4193
TATGTTTCTG	TTTGTAG A	TAT TGT CAG	GAA ATA TAT	AAT TCA AAT	AAT CAA	4244
		Tyr Leu Gln Gln Ile Tyr Asn Ser Asn Asn Gln 135 140 145				
AAO ATT GTT AAC CTG AAA GAG AAG GTA	GCC CAG CTT GAA OCA CAG TGC					4292
Lys Ile Val Asn Leu Lys Gln Lys Val Ala Gln Gln Ala Cys						
		150 155 160				
CAG GAA CCT TGC AAA GAC ACG GTG CAA ATC CAT	GAT ATC ACT GGG AAA G					4341
Gln Gln Pro Cys Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly Lys						
		165 170 175				
GTAACTGATG	AAGGTTATAT	TGGGATTAGG	TTTCATCAAG	TAAATAATGT	AAAGGAGAAA	4401
GTATGTACTG	GAAAGTATAG	GAATAGTITA	GAAAGTGTCT	ACCCATTAGG	TCTAAGAAAT	4461
TCAATGTCTC	AGACCTTTCT	TGAATAAGCTA	AAAAAAGCAG	TTTAAAAAGG	ATGCTGATGT	4521
GAAAAAGTAG	AAAATTATTC	TTGGAAAAATG	AAATAGTTTAC	TACATGTTAA	AAOCTATTTT	4581
TCAAGGCTGG	CACAGTCTTA	CTTGATTTTC	AAACCACAGT	AAAAATGCGT	TCTCTTCTCT	4641
TAG AT TGT CAA GAC ATT GCC AAT	AAG GGA GCT AAA CAG ACG GGG CTT					4688
Asp Cys Gln Asn Asp Ile Ala Asn Lys Gly Ala Lys Gln Ser Gly Leu						
		180 185 190				
TAC TTT ATT AAA CCT CTG AAA GCT	AAC CAG CAA TTC TTA GTC TAC TGT					4736
Tyr Phe Ile Lys Pro Leu Lys Ala Asn Ala Phe Leu Val Tyr Cys						
		195 200 205				
GAA ATC GAT GGG TCT GGA AAT GGA TGG ACT GTG TTT CAG AAG						4778
Gln Ile Asp Gly Ser Gly Asn Gly Trp Thr Val Phe Gln Lys						
		210 215 220				
GTAAATTTTT	CCCCACCATA	TGTATTTAAT	AAATTCCTAC	ATTGTTTTCTG	CCATATGGCA	4838
GATACTTTTT	TAAACACCTT	GTGAACCGTA	GTCTATTTAA	TCTTGTCAAT	AGCCCTAAGG	4898
GGAGGTACTT	TCTGTTACTC	CTATTTACAG	AAAAAGGAAAC	TGAAGCACAC	AAAGTTAAAT	4958
AACCTTGCCCA	AGACCACATA	ACTAATAAGC	AACAAGATCA	GCATTTGAAC	CTAGGCAAGTA	5018
TAGTTTCAGA	GTITGTGACT	TGACTCTATA	TTGTACTGGC	ACTGACTTTT	TGATTTCATG	5078
GTGGCACATA	ATCATAGTAC	CACAGTGAAC	AAATAAAAAAG	AGGAAACTCT	TTTGTCAAGT	5138
AGGTCAAGAC	CTGAGGTTTC	CCATCACAAAG	ATGAGGAAAGC	CCACACCAC	CCCCACCAC	5198
CCCACCACCA	TCACCACCTT	TTACACACCC	AGAGGATACA	CTTGGGCTGC	TCCAAGACAA	5258
GGAACCTGTG	TTGCATCTGC	CACCTGCTGA	TACCCACTAG	GAATCTTGCC	TCTTTTACTT	5318
TCTGTTTACC	TCCCACCACCT	GTATAAAGT	TTTCTACAGG	GGGCGCTCAG	AGGGAATGAA	5378
TGGTGGAAAG	ATTAGTTGCC	AGACACCGAT	TGAGCAATGG	GTTCATCAT	AAATGTAAAG	5438
ATCAAGTAAT	TCCAAGCTAG	GTCTGAGAGT	CGCTAGAGTG	TCTTTTTAAT	ATTACCACTC	5498
ATTTAGAAAT	TATGATGTGC	CAGAAACCCCT	CTTAAGTATT	TCTCTTATAT	TCTCTCTCAT	5558
GATCCTTGCA	GCAACCTTAA	GAAAGTAACCA	TCATTTTTCC	TATTTGATAC	ATGAGGAAAC	5618
TGAGGTAGCT	TGGCCAAAGAT	CACCTAAGTTG	GGAGTTGATA	GAACCAAGTG	TCTGTATTTT	5678
TGACAAAAAT	TTGACAGCAT	TCTCTTTACA	TGCAATGATA	GTCTATTTTC	TCTTTTGTCT	5738

CTTCAGAAATG TATAATTAG AGA CTT GAT GGC AGT GTA GAT TTC AAG AAA AAC	5790
Arg Leu Asp Gly Ser Val Asp Phe Lys Lys Asn	
225 230	
TGG ATT CAA TAT AAA GAA GGA TTT GGA CAT CTG TCT CCT ACT GGC ACA	5838
Trp Ile Glu Tyr Lys Glu Gly Phe Gly His Leu Ser Pro Thr Gly Thr	
235 240 245	
ACA GAA TTT TGG CTG GGA AAT GAG AAG ATT CAT TTG ATA AGC ACA CAG	5886
Thr Glu Phe Trp Leu Gly Asn Glu Lys Ile His Leu Ile Ser Thr Glu	
250 255 260 265	
TCT GCC ATC CCA TAT GCA TTA AGA GTG GAA CTG GAA GAC TGG AAT GGC	5934
Ser Ala Ile Pro Tyr Ala Leu Arg Val Glu Leu Asp Trp Asn Gly	
270 275 280	
AGA ACC AG GTACTGTTTT GAAATGACTT CCAACTTTTT ATTGTAAAGA	5982
Arg Thr Ser	
TTGCCGGAAT TGTGCACTTT CCAACTATCA ATAGACAATG GCAAAATGCAO CCTGACAAAT	6042
GCAAAACAACA CATCCAAGCA CCAATTTTCT CAAGAGTCTG TTTGTTCTTT GGGCAATCCA	6102
AAAAAGTAAA TTCTATTCAO GATGAATCTA AGTGTATTGG TACAATCTAA TTACCCGTGA	6162
ACCATTCAGA GTAAATAGCTA ATTACTGAAC TTTTAACTCA TCCCAAGGAT TGAGCATATAA	6222
ATTATAATTT TAICTAGTCT AAATTACTAT TTCATGAAGC AGGTATTATT ATTAATCCCA	6282
TTTTATAGAT TAACTTGCTC AAGATCAGAT TGCTGATAAG TGATAGAGGT AGAATTCAGA	6342
CTCAAGTATG TTAACCTTTAG AGCCTGTCTT CTAAACAACAT ATCCTGGGTG AAAAAAGCAAT	6402
ACAGCCTCTT CAGACTTCTC AGTGCCTTGA TGCCCATTTA TTCTGTCAAA TCATGAGCTA	6462
CCCTAAAAAT AAACCAAGCTA GCTCTTTTGA TGATCTAGAG GCTTCTTTTT GCTTGAAGATA	6522
TTTGAAGGTT TTAAGCATTO TTACCTAATT AAAATGCAAG AAAATATCCA ACCCTCTTGT	6582
TATGTTTAAO GAATAGTGAA ATATATTGTC TTCAAAACACA TGAACCTTTTT TTTATTGCTT	6642
GGTGTGTTTT TAATCCAAGAA AGTGCTATAG TCAGTAGACC TTCTTCTAGG AAAAGACCTT	6702
CCATTTCCCA GCCACTGGAO ATTAGAAAAA AAGCTAAATA TTTTCTGAAA ATTTCTGTTT	6762
ATTCATTAAO GCCCATCTTT TCCCCCACTC TATAAGAGTO TTGTCCACTT GCACAATTTT	6822
TTCCAAGAAA GAATCTCTCT AACTCTTTCA GCTCACATOC TTTGACACAC ACAAGGAAAG	6882
CTTTGATTGT GTAATGCCCT CAGAAAGCTCT CTTTCTTGGC ACTACCACAC TGATTTGAAG	6942
AAGAAAAATC CTTTGAACAC TAACCTTTCA GGTGCTATGA GTGGCTAATG GAACGTATCC	7002
TCCTTCAATG TTTGTGCAAT AATTAAAGGT CACTCACTGT CAGATACCTT CTGTGATCTA	7062
TGATAATGTO TGTGCAACAC ATAACATTTT AATAAAAGTA GAAATATGAA AATTAGAGTC	7122
ATCTACACAT CTGATTTTGA TCTTAGAATO AAACAAAGCA AAAAGCATCC AAGTGAOTOC	7182
AATTATTAGT TTTCAGAGAT GCTTCAAGAO CTTCAGGCC CATCCGGAO AGTGTAAATG	7242
AGCTGTGAGC TGTTTCACTA ATCTATTGCC TCTTCCAGAA TTTGCAAAAA ACTTCACTCA	7302
ATGAGCAAAAT TTCAAGCTTA AGAAACAAAA TCAAAAAATC CAAGGAAAGCA TCTTCAAGAA	7362
GAAGGAAACTT CTGAGATCCC TGAAGAGGAT CAGCATGTGA TGTTTGTATT TCTTCTTCTT	7422
CAO T ACT GCA GAC TAT GCC ATG TTC AAG GTG GGA CCT GAA GCT GAC	7482
Thr Ala Asp Tyr Ala Met Phe Lys Val Gly Pro Glu Ala Asp	
285 290 295	
AAG TAC CAC CTA ACA TAT OCC TAC TTC GCT GGT GGG GAT GCT GGA GAT	7516
Lys Tyr Asn Leu Thr Tyr Ala Tyr Phe Ala Gly Gly Asp Ala Gly Asp	
300 305 310	
GCC TTT GAT GGC TTT GAT TTT GGC GAT GAT CCT AGT GAC AAG TTT TTC	7564
Ala Phe Asp Gly Phe Asp Phe Gly Asp Asp Pro Ser Asp Lys Phe Phe	
315 320 325 330	

-continued

ACA TCC CAT AAT GGC ATG CAG TTC AOT ACC TGG GAC AAT GAC AAT GAT	7612
Thr Ser His Asn Gly Met Gln Phe Ser Thr Trp Asp Asn Asp Asn Asp	
335 340 345	
AAG TTT GAA GGC AAC TOT GCT GAA CAG GAT GGA TCT GGT TGG TGG ATG	7660
Lys Phe Gln Gly Asn Cys Ala Gln Gln Asp Gly Ser Gly Trp Trp Met	
350 355 360	
AAC AAG TOT CAC GCT GGC CAT CTC AAT GGA GTT TAT TAC CAA G	7703
Asn Lys Cys His Ala Gly His Leu Asn Gly Val Tyr Tyr Gln	
365 370 375	
GTATGTTTT CTTCITTAAG TTCCAAGTTA ATGTATAAGT TATACTATTT TCATAAAAAA	7763
TAATAAATAAG ATATGAAGAA ATGAAGAATA ATTTATAAAG ATAGTAGGGA TTTTATCATG	7823
TTCTTTATTT CAACTAAGTT CTITGAAACT GGAAGTGGAT AATACCAAGT TCATGCTTAA	7883
AATTAGCCCT TCTAAAGAAA TCCACCTGCT GCAAAATATC CAAGTAGTTG GCATTATATG	7943
TGAACCTAT ACCATCATAG CTGGCACTGT ATCTCCTTTA GACATACAAAC	8003
ATAAATGATC TGGATGGATT AACATTACTA CATGGATGCT TGTGGACACA TTAACCTGGC	8063
TCCCATGAG CTITGTGTCA GATACACGCA GTGAACAGGT GTTGGAGGA ACAGAAATAA	8123
GAGAAAGCAA GCACTGGTAA GGGCAAGGGT TTGTGAAGAC TTAGAGAGAA AGACCAATCT	8183
GAGGACAGTA GACACTTATT TTAGGATGGG GGTGGATGTA GGAAGCTATA GTTGTCTATA	8243
AGCTTGGAA GTTGTGGAAC ACTGGTTTCA CTCACCTACC CAGCAATTTT GTGTGGGAAA	8303
GCCTTACCGA TGTAAAGGGA TCCATGTTAC AATAATGGCA TTATTTGAAA ATCCCAATGG	8363
TATTCATAGA ATAAACACAC TATGAAGATA ATCCCACTCA ACAGACTCTC COTTGAGAAA	8423
GGACAGCAAC ACCACCTGG GAAAGCCAAA CAGTCAGACC AGACCTGTTT AGCATCAGTA	8483
GGACTTCCCT ACCATATCTG CTGGGTAGAT GAGTGAAGCC AGTGTTCCAA ACCACTCCGG	8543
GCTTGTAGCA AACCATAATC TCCTCATCTA CCAAGATGAG CAACCTTACC TCTGATGTC	8603
CTAGCCAATC ACCAACTGAG AACCTTGTCA CAGTTTATTT AAAGTAAAGC TTGATTTTTC	8663
ACAATATTTT TAAATGGAG AACATAACT TATCTTGTCA CTCACAAACC ACATAATGAG	8723
AAAGAACTCT AAGGGAAAAAT GCTTGTATCT TGTGACCCGG GGGCCCATGC CAGAGCTGTA	8783
GTTCATGCCA GTTGTGTGCT CTGACAAAGC TTTTACAGAA TTACATGAGA TCTGCTTCCC	8843
TAGGACAGAG AGAAGGCCAA TCAACAGAGG CTGCCTTTA AAATGGAGAC ATAAAAATAC	8903
ATGCCAGAAC CATTTCTCTA AGCTCTCTCA TCAACCAACA AAATTTGTGT TTCAATAAAC	8963
CTGAGTTGAC CTCATCAGGA ATTTTGTGGC TCCTTCTCTT CTAACCTGCC TGAAGAAAAG	9023
TGTTCACAG CAGCTGAGTC GGGGATGGAT AAGCTTAGGG ACAAGGCCCA ATTAAGGAAAC	9083
TTTGGGTTTC TAGCCCTACT AGTAGTGAAT AAATTTAAAG TGTGGATGTT ACTATGAGTC	9143
ACAACACAGA TGTGTGTTAA TAATATGTTT ATTTTATAAA TTGATATTTT AGGAATCTTT	9203
GGAGATATTT TCAATTAAGCA GATAATACTA TAAATTTTAT GTAACTGGCA ATGCACCTTC	9263
TAATAGACAG CTCCTCATAG ACTTGCAGAG GTAAAAAGAT TCCAGAAATA TGATATGATC	9323
ATCTACGACT TGTGTTAG GT GGC ACT TAC TCA AAA GCA TCT ACT CCT AAT	9373
Gly Gly Thr Tyr Ser Lys Ala Ser Thr Pro Asn	
380 385	
GGT TAT GAT AAT GGC ATT ATT TGG GGC ACT TGG AAA ACC GGG TGG TAT	9421
Gly Tyr Asp Asn Gly Ile Ile Trp Ala Thr Trp Lys Thr Arg Trp Tyr	
390 395 400	
TCC ATG AAG AAA ACC ACT ATG AAG ATA ATC CCA TTC AAC AGA CTC ACA	9469
Ser Met Lys Lys Thr Thr Met Lys Ile Ile Pro Phe Asn Arg Leu Thr	
405 410 415	
ATT GGA GAA GGA CAG CAA CAC CAC CTG GGG GGA GCC AAA CAG GTC AGA	9517

-continued

[illegible]

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 453 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(3) SEQUENCE DESCRIPTION: SBO ID NO-6

Met	Ser	Trp	Ser	Leu	His	Pro	Arg	Asp	Leu	Ile	Leu	Tyr	Phe	Tyr	Ala
1				5					10					15	
Leu	Leu	Phe	Leu	Ser	Ser	Thr	Cys	Val	Ala	Tyr	Val	Ala	Thr	Arg	Asp
			20					25					30		
Asn	Cys	Cys	Ile	Leu	Asp	Gln	Arg	Phe	Gly	Ser	Tyr	Cys	Pro	Thr	Thr
		35					40					45			
Cys	Gly	Ile	Ala	Asp	Phe	Leu	Ser	Thr	Tyr	Gln	Thr	Lys	Val	Asp	Lys
	50				55						60				
Asp	Leu	Gln	Ser	Leu	Gln	Asp	Ile	Leu	His	Gln	Val	Gln	Asn	Lys	Thr
	65				70					75				80	
Ser	Gln	Val	Lys	Gln	Leu	Ile	Lys	Ala	Ile	Gln	Leu	Thr	Tyr	Asn	Pro
			85					90					95		
Asp	Gln	Ser	Ser	Lys	Pro	Asn	Met	Ile	Asp	Ala	Ala	Thr	Leu	Lys	Ser
			100					105					110		
Arg	Ile	Met	Leu	Gln	Gln	Ile	Met	Lys	Tyr	Gln	Ala	Ser	Ile	Leu	Thr
		115					120					125			
His	Asp	Ser	Ser	Ile	Arg	Tyr	Leu	Gln	Gln	Ile	Tyr	Asn	Ser	Asn	Asn
	130					135					140				

-continued

Gla Lys Ile Val Asa Leu Lys Gla Lys Val Ala Leu Gla Ala Ala
145 150 155 160

Cys Gla Gla Pro Cys Lys Asp Thr Val Gla Ile His Asp Ile Thr Gly
165 170 175

Lys Asp Cys Gla Asp Ile Ala Asa Lys Gly Ala Lys Gla Ser Gly Leu
180 185 190

Tyr Phe Ile Lys Pro Leu Lys Ala Asa Gla Gla Phe Leu Val Tyr Cys
195 200 205

Glu Ile Asp Gly Ser Gly Asa Gly Trp Thr Val Phe Gla Lys Arg Leu
210 215 220

Asp Gly Ser Val Asp Phe Lys Lys Asa Trp Ile Gla Tyr Lys Glu Gly
225 230 235 240

Phe Gly His Leu Ser Pro Thr Gly Thr Thr Glu Phe Trp Leu Gly Asa
245 250 255

Glu Lys Ile His Leu Ile Ser Thr Gla Ser Ala Ile Pro Tyr Ala Leu
260 265 270

Arg Val Glu Leu Glu Asp Trp Asa Gly Arg Thr Ser Thr Ala Asp Tyr
275 280 285

Ala Met Phe Lys Val Gly Pro Glu Ala Asp Lys Tyr Arg Leu Thr Tyr
290 295 300

Ala Tyr Phe Ala Gly Gly Asp Ala Gly Asp Ala Phe Asp Gly Phe Asp
305 310 315 320

Phe Gly Asp Asp Pro Ser Asp Lys Phe Phe Thr Ser His Asa Gly Met
325 330 335

Gla Phe Ser Thr Trp Asp Asa Asp Asa Asp Lys Phe Glu Gly Asa Cys
340 345 350

Ala Glu Gla Asp Gly Ser Gly Trp Trp Met Asa Lys Cys His Ala Gly
355 360 365

His Leu Asa Gly Val Tyr Tyr Glu Gly Gly Thr Tyr Ser Lys Ala Ser
370 375 380

Thr Pro Asa Gly Tyr Asp Asa Gly Ile Ile Trp Ala Thr Trp Lys Thr
385 390 395 400

Arg Trp Tyr Ser Met Lys Lys Thr Thr Met Lys Ile Ile Pro Phe Asa
405 410 415

Arg Leu Thr Ile Gly Glu Gly Glu Ala His His Leu Gly Gly Ala Lys
420 425 430

Gla Val Arg Pro Glu His Pro Ala Glu Thr Glu Tyr Asp Ser Leu Tyr
435 440 445

Pro Glu Asp Asp Leu
450

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1007 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) ORGANISM SOURCE:

- (B) CLONE: ovine beta- hemoglobin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACGCGTGTGTC ACCTGCAAGT CAACGGATCT CTGTGTCTGT TTTCATGTTA GTACCACT 60

GTITGCTGTC CTGTACCTTT CAGCTACAGT CTGAAGTCAT AAAGCTGTGT ACCTCCAGCT 120

CTGTTCTCTC TCAAGATTGT GTTCTGCTGT TTGGGTCCTT AGTGTCTCCA CACAATTTTT 180

AGAATTGTTT GTTCTAGTTC TGTGAAAAAT GATGCTGATA TTTTGATAAG GATTGCATTO	240
AATCTGTAAA GCTACAGATA TAOTCATTOG GTAGTACAOT CACTTTAACA ATATTAACCTC	300
TTACACATCTO TGAGCATGAT ATATTTTCCC CCTCTATATC ATCTTCAATT CCTCCTATCA	360
GTTCCTTTCA TTGCAATTTT CTGAGTACAG GTCTTACACC TCCTTGTTTA GAGTCATTCC	420
TCAGTATTTT ATTCCCTTGA TACAATTGTO AATGAGGTAA TTTTCTTAOT TTCTCTTTCT	480
GATAGCTCAT TGTAGTGTG TATATAGAAA AGCAACAGAT TTCTATGTAT TAATTTTGTG	540
TCCTGCAACA GATTTCATAT TATTAATTTT GTATCCTGCT ACTTACGGA ATTCACTTAT	600
TAGCTTTTTG GTGACATCTT GAGGATTTTC TGAAGAAAAAT GGCATGATAT GGTAGGACAA	660
GGTGTCTATG CATCTGCAAA CAGTGGCAOT TTTCTTTCTT CCTTCCCAAC CTGGATTTCT	720
TTGATTTCTT TCTGTCTGAG TACGACTAGG ATTCCCAATA CTATACCGAA TAAAGGTGGC	780
AAGAAGTGGC ATCTTTGTCT TATTTTCTG ACCTTAGAGG AAATGCTTTC AGTTTTTCAC	840
CATTAAATAT AATGTTTACT GTGGGCTTGT CATATGTGGC CTTCATTATA TGGAGGTCTA	900
TTCCCTCTAT ACCCACCTTG TTGAGAGTTT TTATCATAAA AGTATGTTGA ATTTGTGCAA	960
AAGTTTTTCC TGACATCTAT GAGATGATTT TTACTCTTCA ATTCAATTAAT GATTTTTIAT	1020
CTTCATTITT TTAATGATTT CCATTCTTCA ATTTGTTAAC GTGGTATATC ACATTGATTO	1080
ATTTGTGAT ACCTTTGTAT CCTTGGATA AACCTCACAT GATCATGAGC TTCAATGTA	1140
TTTTTGAATT CACTTGTCTA ATATTCTGTT GGTGATTTT GATCTCTAT TCATCAATGA	1200
TATTGGCCTA AGAAAAGTTT TGTCTGTTTT TAGTATCAGG GTGATGCTGG CCTCATAGAG	1260
AGAGTTTGA AGCATTTCCT CCTCTTGTAT TTTTCGGAAT AGTTTGAAGT GATAGGTAT	1320
TAACTCTTCT TTAATGTTTT GGGGACTTCC CTGGTGAACC GGTGGTTGAG AATCCGCCCTC	1380
AGGAGTGGG GTTTGTGCCC TGCTCAGGGA ACCATTAAATA AGATCCCAAC TGCTGCAAGC	1440
AACAAAGCCC CAAGCTGCAA CACCTGAGCT GCAACCCCTG CAGTGCCAC AGCCGACGAC	1500
CAGAGAAAAC CCACATACAG CAGGGAAGAC CCAAGCAAC CAGAAAAAGG AGTTTGGTGG	1560
AATACAGCTG TGAAGCCCTC TGTCTCTGGA CTCTGCTTGG AGGGAATTTT TAAAAAATTA	1620
TTGATTCAAT TTCATTACTO GTAACGTGTC TGTTCATATT TTCTATTTCT TCCGGGTTCA	1680
GTCTTGGGAG ATTGTACATG CCTAGGAATG TGTCCGTTTC TTCTAGGTTG TCCATTTTAT	1740
TGACATGCA TGGGAAGACA CAGCACCGAC CAGCGAGACT CATGCTGGCT TCTGGGGGCC	1800
AGGCTGGGGC CCCAAAGCAC ATGGCATCTT AGAGTGTGTG AAAAGCCCACT GACCCCTGCC	1860
AGCCCCACAA TTTCATTCTG AGAAGTGATT CCTTGCTTCT GCATTACAG GCCCCAGATC	1920
TGACCTGCTT CTGAGGAGCA GGGGTTTTGG CAGGACGGGG AGATGCTGAG AGCCGACGGG	1980
GGTCCAGGTC CCTTCCCAAG CCCCCCTGTC TGGGGCAAGC CTTGGGAAAG ATTGCCCAAG	2040
TCCTCCCTGT ACAATGTTCA GTCCCAAGCT CCCCCAGCCA GAGCTGCTTT ATTTCCGTCT	2100
CTCTCTCTGG ATGATATTCT CTGGAAGCTG AAGGTTCTTG AAGTATTGAA TAAGTTTGGC	2160
CTGAAAGGCA TGGTTTGTGG TCAGGTTTCA CAGGAACCTG GAGAGCCCTG CAGCTCAGAC	2220
GTCCCGAGAT TGGTGGCACC CAGATTCTCT AAGCTCCTG GGGAACAGGG CGCTTGTTC	2280
TCCTTGGCTG ACCTCCCTTC TCCTGTCATC ACCCAATTCT GAAAGCAGAG CGGTGCTGGG	2340
GTACAGGCTT CTCGACCTA ACGCCGCTGT CCAAAACCACC CGTGTGTTGG TTCGGGGGGC	2400
TACCTATGGG GAAAGGCTTC TCACTGCAOT GGTGCCCCCC GTCCCTCTG AGATCAGAGG	2460
TCCCAGTCCG GACCTCAAC AGGCCGAGCT CCTTCCAGAG GCTCAAGGGA GGGATCTTGT	2520
CCCCCCCCCT GCTGCTTCCA GCTCTGTTGG CCGACCCCTT GAGCCTGATC TTTAGACGC	2580

CTCAOTCTA	TCTCTGCTC	COTOTTCACA	COCCTTCTCC	CCATOTCCCC	TCCGTOTCCC	2640
COTTTTCTCT	CACAAOGACA	CCOACATTA	GATTAOGCCCC	TOTTTCAOCC	TCACCTOAAAC	2700
AOCTCACATC	TOTAAAOACC	TAOATTCCAA	ACAAOATTCC	AACCTOAAOT	TCCCOTOTGA	2760
TOTGAOTTCT	OGOGCOACAT	CCTTCAACCC	CATCACAOCT	TGCAOTTCTAT	COCAAAACAT	2820
OGAACCTOOG	OTTTATCOTA	AAACCCAOGT	TCTTCATGAA	ACACTGAOCT	TCOAGGCTTO	2880
TTGCAAGAA	TAAAGOTGCT	AATACAGATC	AOGGCAAGGA	CTGAAGCTGO	CTAAGCCTCC	2940
TCTTTCTCAT	ACAAGAAAGG	OGOGCCTOOG	OGCOCCTOGA	OGTCTGCTCC	COTGAOTGAO	3000
CTCTTTCTGO	CTACAOCTCAC	CAACAOTCTC	TCTOOGAAAG	AAACCAAGAG	CCAAGAGACA	3060
AOGCCOAGCT	AOTTTAOGAG	ACCCCTGAAC	CTCCACCCAA	GAOTGTOACC	AOGCAOGCOC	3120
CCCCCTOGAA	AGACCTTACA	OTTCAGOGOG	GAAGAGOGOG	TGACCCOCCA	OGTCCCTGCT	3180
ATCAOGAGAC	ATCCCCGCTA	TCAAGAGATT	CCCCCACCTT	GTCTCCCTTC	CCCTATCCCA	3240
ATACOGCCAC	CCCACCCCTO	TGATOAOGCA	TTTAOTCACT	TAGAAOTGCA	ACTOAAOOGT	3300
TTTGTCTCCC	CTTGTCCAGA	OGCACAAOGC	ACCCACAOGC	TGCTGOGTAC	COACGCCCAT	3360
GTGAATTCAO	CCAAGAGOGC	TGTCTGTGAC	CCTCCCTGCT	COOGCCCCCT	CTGTGTCTGA	3420
CAACACACCC	AOCACCAOCA	TTCCGCTGOC	TGCTGAAGTC	TGCAOGCAOC	TGCTGTOTAOC	3480
CTGAOGGOTO	TGAGOGGAGG	TGTCTGTGGA	GATTTAAAT	GTGAAGOGOG	OGAOGTGOGA	3540
OGTTGOGGCC	TGTGOGGCTO	CCCATCCCA	GTGCTGTGAT	TAOGCCCACT	GTGTCTCAOC	3600
COTGCCCCCO	CCGCAOGGOG	CAOGTCACTT	TCCGCTGCTO	OGGTATTAT	GACTCTGTCT	3660
ATTGCAATTO	CCATTTTTGC	TACCTTAAC	OGGCAOGCAO	TGCTTGAOGA	OGCCTCOATA	3720
CCGACCAOCT	CCTCCCTGGO	AOGTGAACCT	GAACCCCATO	TCACCCCTGOC	CCCAOGCCTOC	3780
AGAGOGGTGO	TGACTGTGGA	GAACCCCTCA	CCCAAGOGCA	COGTCAOGCA	GTGTGAOGGA	3840
GTGTGTGCCC	AAOGCAOGAG	CCACCCCTCA	GAACACACCT	GTCCCCAOTG	CTGTGTCTGA	3900
CCTGTCTCTO	TCTAAGAGOG	TGACCCCGGA	AOTGTCTCTO	GTACTGOGCA	CCAOGCTGGA	3960
CCCAAGAGTCC	AGACACCCAC	CTGTGCCCCC	GTCTGTGOGG	TCTACCAOGA	ACCOTCTAOG	4020
CCCAAGAGOG	ACTTCTGCT	TGOCCTTGOA	TGGAAGAGAG	CTTCTCTATTO	TCTCTGTAOGA	4080
OGAAOGCCAC	CCGOGGCTGO	AOGATGAOGC	AAOTGOGATT	CCGOGAACC	COTGTGCTGO	4140
OGCCCCAOGCC	OGGCTGTGCTO	GGCTGTGATC	CTCTGTATATA	AOGCCCCAAG	CTGTGTGTCT	4200
CAOGCCCTCA	CTCCCTGCAO	AOGTCAOGAG	CACGACCCCA	OGGATATCCC	TGCAOGCCATG	4260
AAOTGTCTCC	TGCTTGTGCT	OGGCTGTGOC	CTGCTGTGTG	OGTGTCAOGC	CATCATGCTC	4320
ACCCAGAGCA	TGAAAGOGCT	OGACATCCAG	AAOGTTCGAO	OGTTGOGCOC	GTGTGTGAOT	4380
TGCAOGGCGO	OGAOGGAGOC	TGOGCCTCAO	AGAGCCCAAG	GAOGGCTGTGA	COTGTGOGTT	4440
CCATCAOTCA	OGTAOGGCA	CCTGACAAAT	CCCGGCTGOG	OGAOGTTCA	CCAOGGCTTC	4500
ACTGTCTTGC	ATTCTGTGAOG	CTGGAAGOGC	AAOGTCCAGG	TGTGTGCAOG	GTGTGTGTCT	4560
CCTGOGGCGO	CTCTGTGOGG	AOGAGACOGC	GTCTTGTCTC	AOTGTCTGTG	OGGCGCTGAT	4620
TTCTCTTCTC	TGTGAOGGCA	CCAOGGCTGC	TGGAAGACAG	CCTGTGCTGC	CAOGTTTACA	4680
COACCTTTGT	CATCTCTTTA	AAOGCCATGT	CTCCAGAGATC	ATGTGTGTA	GTCTGTGOGG	4740
TTAGTGOAGC	ACAOTTCAOC	CCCTAAAGAG	GTCTGTCTGC	CCCTCAAAAT	TTCCCCACCT	4800
CGAOGGATOT	CTCCCCAAGA	CTCAAAOTGT	GTACATGTG	OGGOGGCTCA	TCTGOGTCCC	4860
TCTTTGOGTT	CAOTGTGAOT	CTGOGGAGAG	CATTCCCCAG	OGTGTGAAGT	TGOGGOGGAGT	4920
ATCTCAOGGC	TGCCCCAOGC	OGGOTGOGAC	AGAGAGCCCA	CTGTGOGGCT	OGGOGGCCCC	4980

TCCCAACCCC	AGAOTGCAAC	TCAAAGOTCCC	TCTCCAAGT	GC0000ACTT	GCACCTCCTT	5040
GGCTAT0000	GCCA0C0ACA	TCTCCCT0CT	00AT0CCCA0	AGT0CCCCCC	T0A0A0T0TA	5100
CGT00A00A0	CT0A00CCCA	CCCC0A000	CAACCT00A0	ATCCT0CT0C	AGA0A0T0T0	5160
GGC0TCTCTC	CCCAACAT00	AACCCCCACT	CCCCA000CT	0T00ACCCCC	C000000T00	5220
0GT0CA00A0	00ACCA0000	CCCA0000CT	000AA0A000	CTCAG0ATTT	ACT00TACCC	5280
GGC0CTCCAC	CCAA00CT0C	CCACCA000	CTTTTTTTTT	TTTTAAACTT	TTATTAATTT	5340
GATGCTTCA0	AACATCATCA	AACAAAT0AA	CATAAAACAT	TCATTTTT0T	TTACTT00AA	5400
0000AGATAA	AATCCTCT0A	AGT00AAA0T	CATAGCAAA0	ATACATACAA	T0A00CA00T	5460
ATTCT0AATT	CCCT0TTA0T	CT0A00ATTA	CAA0TGTATT	T0AGCAACA0	AGA0ACATTT	5520
TCATCATTTC	TA0TCT0AAC	ACCTCA0TAT	CTAAAAT0AA	CAA0A0T0CC	T00AAA00AA	5580
0CA0T0T000	GATA00CCCC	T0T00A00CT	0CT000A00C	AGCA0ACCT0	00TCTTC000	5640
CTCAA0CA0T	TCC0CTACCC	A0CCCC0TCC	ACCTCA0AC0	0000TCA000	T0CA00A0A0	5700
A0CT00AT00	0T0T00000C	AG0AT0000	ACCT0A0CCC	CA000CT0CC	TTTT00000T	5760
0CCT0T00TC	AA00CTCTCC	CT0ACCTTTT	CTCTCT00CT	TCATCT0ACT	TCTCCT00CC	5820
CATCCACCC0	0TCCCCCT0T	0CCT0A00T0	ACA0T0A0T0	C0CC0A00CT	A0TT00CCAA	5880
CT00CTCCTA	T0CCCAT0CC	ACCCCCCTCC	A0CCCCCT0	00CCA0CTTC	T0CCCCCT00C	5940
CCTCA0TTCA	TCT0AT0AA	AAT00TCCAT	0CCAAT00CT	CA0AAA0CA0	CT0TCTTTCA	6000
000A0AAC00	CGA0T0T0CT	CAGAA0AA0A	TTATT0CA0A	AAAAACCAA0	ATCCCT0C00	6060
T0TTCAA0AT	CGAT00T0A0	TCC000TCC0	T00000ACAC	CCACCACCCC	C0CCCC0000	6120
0ACT0T00AC	A00TTCAG00	00CT00C0TC	000CCCC000	AT0CTAA000	ACT00T00T0	6180
AT0A00ACAC	T0CCTT0ACA	CCT0CTTTCAC	TT0CCTCCCC	T0CCACCT0C	C000000CCTT	6240
0000C00T00	CCAT000CA0	0TCCC00CT0	0C000CTAAC	CCACCA000T	0ACACCC0A0	6300
CTCTCTTT0C	T000000C00	0C00T0CTCT	000CCCCA0	0CT0A0CTCA	00A00TACCT	6360
0T0CCCCTCC	A0000TAACC	0A0A0CC0TT	0CCCACTCCA	0000CCCCA0	T0CCCCAC0A	6420
CCCCA0CCCC	CTCCACA0CT	CCTTCATCTC	CT00A0ACAA	ACTCT0TCC0	CCCTC0CTCA	6480
TTCACTT0TT	C0TCTCTAAAT	CC0A0AT0AT	AAA0CTTCA0	0000000T0T	000TTCATCA	6540
A000CT0CCC	TTC000000	CA0CCT000C	CACATCT0CC	CTT00CCCCC	TCA00ACTCA	6600
CTCT0ACT00	A00CCCC0CA	CT0ACT0AC0	CCA000T0CC	CA0CCCCA00	CTCT00C0C	6660
CATCCA0CT0	CACCT000TTT	000T0CT00T	CCT0CCCCCA	A0CT0CCCC0	ACACCCAC00	6720
CA0CC0000C	T0CCCCACT0	CCTC00TCA0	00T0A0CCCC	AGCT0CCCCC	0CTCA000CT	6780
T0CCCC0ACA	AT0ACCCCAAT	CCTCA00AC0	CACCCCCCTT	CCCTT0CT00	0CA0T0TCCA	6840
0CCCCACCC0	AGATC00000	AA0CCCCATT	TCTT0ACAAAC	TCCA0TCCCT	00000A0000	6900
0CCTCA0ACT	0A0T00T0A0	T0TTC0CAA0	TCA00A00T	00T00A000T	CCT00C00AT	6960
CCA0A0TT0A	CA0T0A000C	TTCCT000CC	CCAT0C0CCT	00CA0T00CA	0CA000AA0A	7020
00AA0CACCA	TTTCA0000T	00000AT0CC	AG00C00CTC	CCCACCC00T	CTT00CC000	7080
T00T0ACCCC	00000A0CCC	C0CT00TCT0	00A000T0CT	00000CT0AC	TAGCAACCCC	7140
TCCCCCCCC0	TT00AACTCA	CTTTTCTCCC	0TCTT0ACCC	C0TCCA0CCT	T0AAT0A0AA	7200
CAAA0TCCTT	0T0CT00ACA	CC0ACTACAA	AAA0TACCT0	CTCTTCT0CA	T00AAAA0A0	7260
T0CT0A00CC	0A0CAAA0CC	T00CCT0CCA	0T0CCT000T	000T0CCAAC	CT00CT0CC	7320
CA000A0ACC	A0CT0C0T00	TCCTT0CT0C	AAC00000T	000000T000	A0CTT0ATCC	7380

-continued

```

ACCGTAGCCT TCCAGGTTTC TTCTGTCCAC AGAATTCTCC AAGGCAAGAA TACTGGAATG 9840
GGTAGCCATT TCTCTCTCCA GGGGATCTCT CCGACCCAAG GATTGAACCG GCATCTCTCT 9900
TATTGGCAAG TGGATTCTTT ACCACTGTGC CACCAAGGAA GCCCGTGTTA CTCTCTATGT 9960
CCCACCTAAT TACCAAGGCT GCTCCAAGAA AAGGCCCTGT TGCCCTCTGA GCTTCCCGGC 10020
CTGCAAGGAG TGGTGGGGGT AGACTGTGAC CTGGGAACAC CCTCCCGCTT CAGGACTCCC 10080
GGGCCACGTC ACCCACAGTC CTGCAAGCAG CCGGTAAGCT CTGCTCTTCA AGGCTCATT 10140
TCTTAAAAA AACTGAGGCT CTATTTTGTG ACTTCGCTGC CGTAACCTCT GAACATCCAG 10200
TGCATGAGC AGGACCTCTT CCCCAGGCTT CAGGGGCTTC AAGGAGCCAG CCTTCACCTA 10260
TGAATCACCA GACACTCGGG GGTGGCCCCG CCTTCAGGCT GCTCACAGTC TTCCCATCGT 10320
CCTGATCAAA GAGCAAGAAC AATGACTTCT TAGGAGCAAG CAGACACCCA CAGGACACTG 10380
AGGTTACCCA GAGCTGAGCT GTCTTTTGA ACCTAAAGAC ACACAGCTCT CGAAGGTTTT 10440
CTCTTTAATC TGGATTAAAG GCCTACTTGC CCTCAAGAG GGAAGACAGT CCTGCTATC 10500
CCCAAGGAGC CCACCTGGGT GCATCCGAG CCACCTTAAT TTATCTGACC GCACCCCTGA 10560
ATTAATCGGT CCAAACTGGA CAAAACCTT GGTGGGAAGT TTATCCCAAG AGGCTCAAC 10620
CATCCTGCTT TGACCACCTT GCATCTTTT TTCTTTTATG TOTATGATG TATATATATA 10680
TATATATTTT TTTTTTTTCT ATTTTTTGGC TGCTCTGGCT GTTCGTTGCA GTTCGGTGGC 10740
CAGGCTTCTC TCTAGTTTCT CTCTAGTCTT CTCTTATCAC AAGGCAAGCT CTAGACGATC 10800
GACGCGCT 10860

```

(2) INFORMATION FOR SEQ ID NO:8:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AATTCCGATC GACGCGTGA CGATATATCT TAGACGATCG ACGGCTA

47

(2) INFORMATION FOR SEQ ID NO:9:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) IMMEDIATE SOURCE:

- (B) CLONE: BLOAMP3

(2) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGGATCCCCC GCGGCTGCT CTGG

24

(2) INFORMATION FOR SEQ ID NO:10:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) IMMEDIATE SOURCE:

- (B) CLONE: BLOAMP4

(2) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACGCGTCAT CCTCTGTGAG CCAAG

24

-continued

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (v i i) IMMEDIATE SOURCE:
 (B) CLONE: ZC6839

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTACGTAOT

1 0

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (v i i) IMMEDIATE SOURCE:
 (B) CLONE: ZC6632

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCACGCGGAT CCTACGTACC TOCAGCCATO TTTCCATGA GG

4 2

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (v i i) IMMEDIATE SOURCE:
 (B) CLONE: ZC6627

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGGCTTCGG CAAAGCTTCAG G

2 1

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (v i i) IMMEDIATE SOURCE:
 (B) CLONE: ZC6521

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCCAAAGACT TACTTCCCTC TAGA

2 4

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (v i i) IMMEDIATE SOURCE:
 (B) CLONE: ZC6500

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCATGAACOT CCGTGTGTGG TTGTGCTACC

3 0

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6519

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ACCACGCCAC GTTCATGCTC TAAAAACCGT

3 0

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6518

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTGCGGGAT CCTACGTACT AGGGGGACAG GGAAGG

3 6

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6629

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGACGCGAAT TCTACGTACC TGCAGCCATO AAAAGGATGG TTCTC

4 5

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6630

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGACGCGAAT TCTACGTACC TGCAGCCATO AAACATCTAT TATTO

4 5

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6625

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTGAGATTTT CAGATCTTGT C

2 1

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6636

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGAATTACT GTGGCCTACC A

21

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6634

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTGCGGAAT TCTACOTACT ATTGCTGTGG GAA

33

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: ZC6514

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGACGCGGAT CCTACOTACC TGCAGCCATG AGTTGGTCCT TCAC

45

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: sc6517

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTCTCTGTA GCAACATACT A

21

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: sc6516

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGGTTTCTAG CCCTACTAAT AG

22

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(v i i) IMMEDIATE SOURCE:

- (B) CLONE: uc5515

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGGTTTCTAG CCCTACTAGT AG

22

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AAGCTACGCG TCGATCCTCT AAGATATATC CTCGACCGGT CGATCGG

47